

Auto-antonym の多カーネルオンライン意味変化分析

岩本 蘭¹ 湯川 正裕^{1,2}

1. 慶應義塾大学 2. 理研 AIP

iwamoto@ykw.elec.keio.ac.jp, yukawa@elec.keio.ac.jp

1 はじめに

近年 SNS の普及により誰もが気軽に文章を投稿できるようになり、たくさんのお新語が発生し単語の意味も急速に変化している。意味変遷の結果、反対の意味の複数の語義を持つようになった単語を Auto-antonym といい、例えば「clip」や「dust」、「やばい」等が挙げられる。Auto-antonym を含む文章は作られた年代によって意味が異なるため、既存の分散表現を用いたバッチ学習法では意味解析が難しい。なぜなら (1) 語義曖昧性を持つ多義語が複数の特徴ベクトルを持たない、(2) 単語が途中から複数の意味を持つように変化する場合、バッチ学習ではそれぞれの意味の時系列的な分析が困難である、という問題が生じるからである。

本論文では非線形適応学習法である多カーネル適応フィルタ [1] に基づくオンライン分類器を用いた意味変化分析を検討し、ある単語の新しい意味が産生し元の意味と共存していく過程を可視化する。Auto-antonym を含む文章を時系列順にオンライン極性分類 (positive, negative) することによって単語の意味変遷を自動検出することを目的としている。Auto-antonym の一例として「sick」という単語に着目し、本来の意味である「ill」、比較的新しい「cool」、また文脈だけでは意味を判断できない場合の 3 グループそれぞれの通時的変化を示す。特徴ベクトルではなく分類器に焦点をあて、複数の意味を同時に持つ単語や短時間で意味が変遷する単語に対して分類境界を適応的に更新するアプローチの有用性を検討する。

2 関連研究

初めに単語分散表現と意味変化の研究の歴史について述べる。word2vec [2] を始めとした単語分散表現の登場により特徴空間内で単語の意味演算が可能になり、次元が大幅に削減されたことで大規模処理が容易になっている。また、分散表現モデルをオンライン処理で更新する研究 [3] も行われている。

年代ごとに分散表現を学習した研究 [4] では、ある単語と共起する単語の変化を調べることで意味変遷を捉えており、大量のデータが既に存在し意味変遷が長期間にわたるときに有用な方法である。しかし学習に用いる年代の幅によっては意味の変遷が捉えられず、またその年代の幅の中ではデータを学習するとき時間的順序は考慮されていない。そこで単語の意味変化

の速さを時間微分として考える連続的な分散表現モデル [5] が考案されている。しかし様々な単語の変遷を一度に捉えることは依然として難しく、また事前に分散表現を学習するために大量のデータが必要となる。

本論文では分散表現に代表される特徴ベクトルではなく非線形オンライン分類器を用いて意味変化を分析する。以下にオンライン学習の既存研究について述べる。訓練データを正しく分類できなかった時のみ重みを更新する Passive Aggressive アルゴリズム [6] は Gmail の優先トレイ分類に用いられる。オンライン線形 SVM 学習手法である Pegasos [7] は最適解に近い解への収束が早く大規模データ分析に適している。しかし非線形カーネルでは性能が低下することが知られている。他に重みベクトルが正規分布から生成されていると仮定する Confidence Weighted Algorithm [8] や、CW に頑健性を加えた Adaptive Regularization of Weight Vectors [9] 等があげられる。提案手法は複数のカーネル (ガウスカーネル) を組み合わせ、分類境界の局所的な形状に合わせて過学習を防ぎながら適応的に推定を行う。低次元の特徴ベクトルの研究の発展により特徴空間が複雑化するに伴い、複雑な分類境界を逐次的に推定できる分類アルゴリズムの必要性が高まっている。

3 実験

3.1 設定

Auto-antonym (極性が反対の複数の語義を持つ単語) の意味変遷をオンライン学習で分析する。Auto-antonym の例として英単語「sick」を用いる。「sick」は本来「病気の」、「うんざりして」、「気味の悪い」という否定的な意味を持つが、最近では若者の間で「cool」、「excellent」の意味でも用いられる。コーパス検索ツール Sketch Engine¹ で文例を取得する。使用したコーパスを表 1 に示す。逐次的な意味変化分析のラベル付きデータセットを取得することが難しいため、VADER Sentiment Analysis [10] を用いてラベル付けを行い感情極性スコアが強いものをラベル付きデータとして採用する。VADER では人力でのアノテーションに品質を担保する仕組みをつけ、文中の but や感情に関連する単語の大文字表記、感嘆符の数等をもとに

¹<https://www.sketchengine.eu/>

表 1: コーパス

コーパス名	トークン数	用途	データ数	年
British National Corpus (BNC)	112,345,722	Train	1913	1960-1994
Timestamped web corpus 2005-2014 (Feed)	640,820,898	Train	3874	2006-2014
Timestamped JSI web corpus 2014-2016 English	21,336,894,049	Train	3235	2014-2016
		Test	3314	

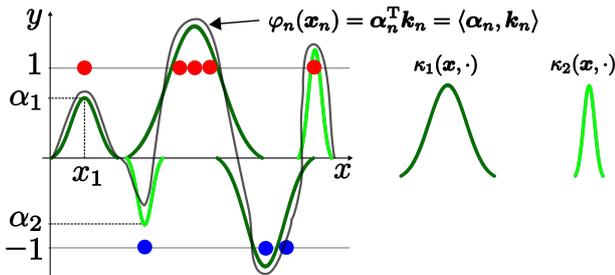


図 1: 多カーネルオンライン分類器

感情極性を判断している。単語分散表現の学習済みベクトル LexVec trained English Wikipedia 2015 and NewsCrawl datasets [11] を用いて文書ベクトルを作成する。訓練データを時系列順に並べ分類器に学習させることにより通時的な意味変遷を分析する。

3.2 意味変化分析

実験には基底ベクトルをオンラインで構築する非線形適応学習法である多カーネル適応フィルタ [1] を拡張した分類器 [12] を用いる。分類境界を非線形関数として考え幅の異なる複数のカーネルを用いてオンライン学習を行う手法で、詳細については Appendix で述べる。図 1 は提案法の分類境界推定手法を表しており、(i) 辞書更新と (ii) 係数更新の二つのステップを繰り返し分類境界を推定する。

分類器はその時点までの訓練結果をふまえて今入ってきた訓練データを適切に分類するように学習しているので、テストデータは分類器の訓練が終わった時の年代である 2014-2016 年の文章を用いる。まずオンライン学習手法のライブラリである LIBOL [13] を用いて既存のオンライン学習法と提案法を比較する。

テストデータの学習曲線を図 2 に示す。提案法の精度は 81.6204, 比較手法では Second Order Perceptron, Perceptron, Passive Aggressive, Confidence-Weighted learning アルゴリズムの順に 57.574, 77.188, 81.563, 80.658 の精度を記録している。我々の提案した分類器は既存のオンライン分類器と同等の性能を出している。

次に sick の意味変遷を調べる。SNS では短文の投稿が多く、またある文章が前のどの投稿に対して書かれたのかを分析することが難しいため、単体では極性

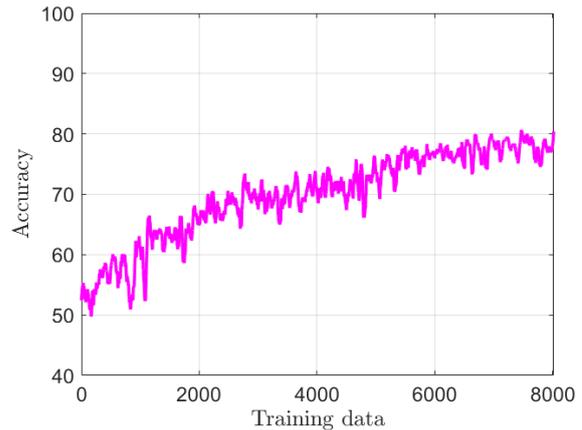


図 2: 学習曲線

が判定できない文章の予測をすることは重要なタスクである。今回は Auto-antonym の時系列での極性分析を行うことにより意味の変遷を可視化する。Oxford dictionary², Collins dictionary³などから「sick」を含む文章を意味別に 3 グループ、20-30 文ずつ選び意味変化の検証データとする。文章の意味ラベルが正確なものを用いるために辞書の例文を検証データとして採用している。「sick」の分類とグループごとの文章の一部を表 2 に示す。3.2 節で学習した分類器を用いて時系列での推定値の変化を調べる。時系列順に並んだ訓練データが入ってくるごとに例文の予測スコアを計算する。つまり、検証データとして用意している例文がその年代においてどのような極性を持っていると評価されるのかを分析する。図 3 は A, B, C のグループごとの文章の極性スコアの平均をとり、感情極性の変化を示したものである。否定的な意味を持つグループ A は訓練データが増えても (時間が経っても) ずっと推定値はマイナスのままである。肯定的な意味を持つ B のグループの文章は、初めは否定的な意味だと判断されているが時間が経つにつれ肯定的な意味と判断されるようになっていく。これは、1980-2000 年にはまだ肯定的な意味で「sick」が使われることは少なかったという現状を反映した推定結果だと考えられる。つまりオンライン学習ではバッチ学習ではとらえられない意味の変遷を分析できることが分かる。

²<https://en.oxforddictionaries.com/>

³<https://www.collinsdictionary.com/dictionary/english>

表 2: sick の意味変遷の例文

	例文
Group A: 病気, うんざりして, 気味の悪い (negative)	I'm sick at not getting that job. My father has been off sick for a long time. Looking at that car accident made me feel sick.
Group B: cool, excellent (positive)	Dude, those shoes are sick! It was a sick party and there were tons of cool people there. Your new car is sick.
Group C: どちらの極性にも解釈できる	You are sick. That's sick I think you look sick.

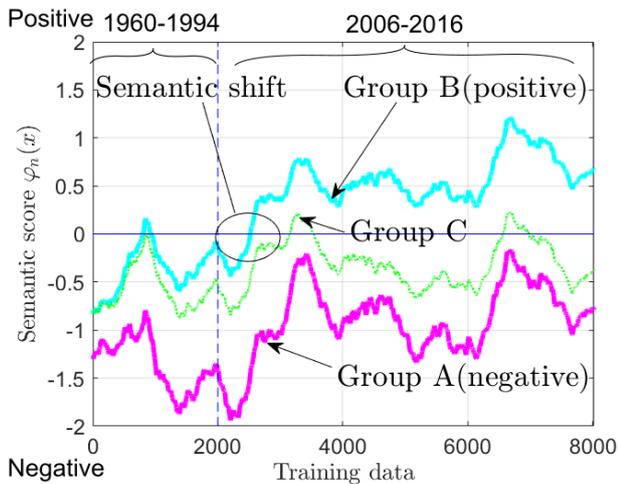


図 3: sick の意味変遷分類

4 おわりに

本論文では Auto-antonym の一例として「sick」を取り上げ、多カーネル適応フィルタを用いたオンライン分類器で意味変化を分析した。時間と共に本来の意味とは反対の意味を獲得し現在ではどちらの意味も使用される単語の複数の意味変遷を、オンライン学習により追跡できることを示した。またタスクごとに最適化した特徴ベクトルではなく訓練済み単語ベクトルを用いることで特徴ベクトル作成の労力を割かず分析を行うことができた。

分散表現を用いた既存研究では「gay」のようにある単語の意味が全く別の意味に変遷する様子を捉えることはできていた。しかし分散表現では基本的に1単語、1文章につき1つの特徴ベクトルを持つため2つ以上の意味が同時に存在する状況で両方の意味変化を分析することはできていなかった。提案法ではオンライン学習により時系列順に並べた訓練データを学習していくにつれて肯定的な意味で「sick」が用いられている文章が新しい意味に認識されていく過程を示した。

Auto-antonym のように複数の意味を持つ単語の意味変遷を推定するためのラベル付きデータセットの構築がこれからの課題であり、年代ごとに均一な数の訓練データを使用することにより更なる検証を行いたい。

謝辞

研究の初期段階で有益なご助言をいただいた慶應義塾大学の小原京子教授に心から感謝申し上げる。本研究は科研費の助成 (18H01446, 15H02757) による。

Appendix

実数全体の集合を \mathbb{R} 、非負実数全体の集合を \mathbb{N} 、正実数全体の集合を \mathbb{N}^* とする。ベクトル $\mathbf{a}, \mathbf{b} \in \mathbb{R}^p, p \in \mathbb{N}^*$ の内積とノルムを $\langle \mathbf{a}, \mathbf{b} \rangle := \mathbf{a}^T \mathbf{b}, \|\mathbf{a}\| := \sqrt{\langle \mathbf{a}, \mathbf{a} \rangle}$ と定義する。ここで $(\cdot)^T$ は転置を表す。

提案法の分類器について述べる。2クラス分類問題で入力ベクトル $(\mathbf{x}_n)_{n \in \mathbb{N}} \subset \mathcal{U}$ とラベル $(y_n)_{n \in \mathbb{N}} \subset \{+1, -1\}$ が逐次的に観測される状況を考える。入力空間は $\mathcal{U} \subset \mathbb{R}^L, L \in \mathbb{N}^*$ で表す。分類境界の非線形関数 $\varphi: \mathcal{U} \rightarrow \mathbb{R}$ を求めることが目的である。出力は実数なので $\varphi(\mathbf{x}_n) > 0$ のとき $y_n = +1, \varphi(\mathbf{x}_n) < 0$ のとき $y_n = -1$ と分類する。入力 \mathbf{x}_n の再生核 $\kappa(x_n, \cdot)$ を辞書 $\mathcal{D}_{q,n} := \{\kappa_q(\cdot, \mathbf{x}_j)\}_{j \in \mathcal{J}_{q,n}}, n \in \mathbb{N}$ に追加し、辞書が張る空間の中で φ を推定する。時刻 n における入力データ \mathbf{x} に対する φ の出力は、係数 $\alpha_{j,n}^{(q)} \in \mathbb{R}, \mathcal{J}_{q,n} \subset \{0, 1, 2, \dots, n\}$ と幅の異なるガウスカーネル

$$\kappa_q(\mathbf{x}, \mathbf{y}) := \exp\left(-\frac{\|\mathbf{x} - \mathbf{y}\|^2}{2\sigma_q^2}\right), q \in \mathcal{Q} := \{1, 2, \dots, Q\} \quad (1)$$

を用いて

$$\varphi_n(\mathbf{x}) = \sum_{q \in \mathcal{Q}} \sum_{j \in \mathcal{J}_{q,n}} \alpha_{j,n}^{(q)} \kappa_q(\mathbf{x}, \mathbf{x}_j), n \in \mathbb{N}, \mathbf{x} \in \mathcal{U} \quad (2)$$

と書ける。

辞書更新

一番幅の広いカーネル κ_1 を用いて大まかな分類境界を定め、次に幅の狭いカーネル κ_2 を用いて局所的に複雑な分類境界を推定する。新しい入力 \mathbf{x}_n が到着したとき、まずヒンジ誤差 [14] を用いた誤差条件

$$\max\{1 - y_n \varphi_n(\mathbf{x}_n), 0\} > \epsilon, \epsilon \in (0, 1) \quad (3)$$

を考える。入力 \mathbf{x}_n が正しく分類され誤差が小さい場合には辞書を更新しない。次に幅の広いカーネルから順にコヒーレンス条件 [15]

$$\max_{j \in \mathcal{J}_{q,n}} \left| \frac{\kappa_q(\mathbf{x}_n, \mathbf{x}_j)}{\sqrt{\kappa_q(\mathbf{x}_n, \mathbf{x}_n) \kappa_q(\mathbf{x}_j, \mathbf{x}_j)}} \right| \leq \delta, \delta \in (0, 1) \quad (4)$$

を適用し、条件を満たす $\kappa_q(\mathbf{x}_n, \cdot)$ のみを辞書に追加することで辞書の冗長性を減らし過学習を抑える。最後にシュリンク条件

$$|\alpha_{j,n}^{(q)}| \leq \gamma \max\{|\alpha_{i,n}^{(p)}|\}_{i \in \mathcal{J}_{q,n}, p \in \mathcal{Q}}, \gamma \in (0, 1) \quad (5)$$

を用いて係数が小さく推定に貢献していない辞書点を辞書から削除することにより辞書サイズを小さくする。

係数更新

係数 $\alpha_{j,n}^{(q)}$ の初期値を 0 とする。推定関数は

$$\varphi_n(\mathbf{x}_n) = \boldsymbol{\alpha}_n^T \mathbf{k}_n = \langle \boldsymbol{\alpha}_n, \mathbf{k}_n \rangle \quad (6)$$

で与えられる。誤差関数としてシグモイド関数を用いた。係数ベクトルの更新式はステップサイズ $\eta \in [0, 2]$ と正則化パラメータ $\rho > 0$ 、シグモイド関数のパラメータ $b \in [0, 1]$ を用いて

$$\boldsymbol{\alpha}_{n+1} := \boldsymbol{\alpha}_n + \eta \frac{\frac{1}{y_n} \log \frac{1-b}{b} - \langle \boldsymbol{\alpha}_n, \mathbf{k}_n \rangle}{\|\mathbf{k}_n\|^2 + \rho} \mathbf{k}_n \quad (7)$$

で与えられる。

実験で用いた数値について述べる。本実験では幅の異なる 2 個のカーネルを用いて分類境界を推定した。パラメータチューニングとしてカーネルスケールパラメータ $\sigma^2 \in \{6, 8, 10, 15, 20\}$ 、コヒーレンスパラメータ $\delta \in \{0.2, 0.3, 0.4, 0.5\}$ 、ステップサイズ $\eta \in \{0.01, 0.21, 0.41, 0.61, 0.81, 1.01\}$ 、誤差条件 $\gamma \in \{0, 0.005, 0.01\}$ 、正則化パラメータ $\rho \in \{0.01, 0.21, 0.41, 0.61, 0.81\}$ をグリッドサーチで最適化し、最終的なパラメータとして $\delta = 0.3, \epsilon = 0.05, \eta = 0.81, \rho = 1.01, \sigma_1^2 = 400, \sigma_2^2 = 36, \gamma = 0.05$ を用いた。辞書サイズは $D_{1,n} = 41, D_{2,n} = 2320$ だった。

参考文献

- [1] M. Yukawa, “Multikernel Adaptive Filtering,” *IEEE Transactions on Signal Processing*, vol. 60, no. 9, pp. 4672–4682, Sept. 2012.
- [2] T. Mikolov, I. Sutskever, K. Chen, G. S. Corrado, and J. Dean, “Distributed Representations of Words and Phrases and their Compositionality,” in *Advances in NIPS*, pp. 3111–3119, 2013.
- [3] N. Kaji and H. Kobayashi, “Incremental Skip-gram Model with Negative Sampling,” in *Proceedings of the 2017 Conference on EMNLP*, 2017, pp. 363–371.
- [4] W. L. Hamilton, J. Leskovec, and D. Jurafsky, “Diachronic word embeddings reveal statistical laws of semantic change,” in *Proceedings of the 54th Annual Meeting of the ACL*, 2016, pp. 1489–1501.
- [5] A. Rosenfeld and K. Erk, “Deep Neural Models of Semantic Shift,” in *Proceedings of the 2018 Conference of the NAACL*, 2018, pp. 474–484.
- [6] K. Crammer, O. Dekel, J. Keshet, S. Shalev-Shwartz, and Y. Singer, “Online Passive-Aggressive Algorithms,” *JMLR*, vol. 7, pp. 551–585, Jan. 2006.
- [7] S. Shalev-Shwartz, Y. Singer, and N. Srebro, “Pegasos: Primal Estimated sub-GrAdient Solver for SVM,” in *Proceedings of the 24th International Conference on Machine Learning*, 2007, ICML ’07, pp. 807–814, ACM.
- [8] K. Crammer, M. Dredze, and A. Kulesza, “Multi-Class Confidence Weighted Algorithms,” in *Proceedings of the 2009 Conference on EMNLP*, 2009, pp. 496–504.
- [9] K. Crammer, A. Kulesza, and M. Dredze, “Adaptive regularization of weight vectors,” in *Advances in NIPS*, 2009, pp. 414–422.
- [10] C. J. Hutto and E. Gilbert, “VADER: A Parsimonious Rule-Based Model for Sentiment Analysis of Social Media Text,” in *Proceedings of ICWSM 2014*, 2014.
- [11] A. Salle, A. Villavicencio, and M. Idiart, “Matrix Factorization using Window Sampling and Negative Sampling for Improved Word Representations,” in *Proceedings of the 54th Annual Meeting of the ACL*, 2016, vol. 2, pp. 419–424.
- [12] R. Iwamoto and M. Yukawa, “Sentiment Analysis in Twitter Based on Multikernel Adaptive Filtering,” in *IEICE SIP SYMPOSIUM*, 2018.
- [13] S. C. Hoi, J. Wang, and P. Zhao, “LIBOL: A Library for Online Learning Algorithms,” *JMLR*, vol. 15, pp. 495–499, 2014.
- [14] C. M. Bishop, *Pattern Recognition and Machine Learning (Information Science and Statistics)*, Springer-Verlag, 2006.
- [15] C. Richard, J. C. M. Bermudez, and P. Honeine, “Online Prediction of Time Series Data With Kernels,” *IEEE Transactions on Signal Processing*, vol. 57, no. 3, pp. 1058–1067, Mar. 2009.