

プロフィール情報を用いたテキストの視点分類

吉田 稔^{1,a)} 松本 和幸^{1,b)} 北 研二^{1,c)}

概要: 同一トピックに関する Twitter への投稿を、視点によって分類する手法を提案する。提案手法は、ユーザーのプロフィール情報を援用し、各投稿内容の分散表現を学習することで、類似する視点からの投稿をまとめる。

1. はじめに

あるテーマについて、与えられたテキストが賛成の立場か反対の立場かを分類するスタンス分類 (Stance Classification) の研究は、近年、SNS 上で様々なユーザーが様々な議論を行うことが一般的になるにつれ、活発に研究が行われている [10]。通常、スタンス分類においては、あるテーマについて、各投稿が「賛成」か「反対」かを分類するが、本研究では、より広い分類タスクとして、ユーザーの視点によりテキストを分類するタスクを提案する。

同一単語に関する文書を意味ごとに分類するタスクは、曖昧性解消の問題として広く扱われているが、本研究の想定する状況は、単語の意味そのものに曖昧性は無いものの、それに関する言及が、ユーザーの種類によって異なる性質になるという状況である。例えば、「徳島」という単語を表す場合、「徳島」という単語自体に曖昧性は無くとも、「県内在住のユーザー」と「県外在住のユーザー」で興味のある情報は異なり、「徳島」に関する言及も異なった傾向になると考えられる。(例えば、県外のユーザーは旅行に関する言及、県内のユーザーは生活情報に関する言及が多くなることが予想される。) 投稿をこのようなユーザーの視点によって分類することができれば、関係の無い投稿を除外する、自分と違う視点の投稿を俯瞰する、等、SNS のユーザーインターフェースを大きく改善させることが期待できる。また、本タスクは、従来のトピック抽出・アスペクト抽出の研究等のように特定のトピックやアスペクトを抽出するのではなく、トピックやアスペクトの傾向も含めた発言者の志向を分類することに主眼をおく。

本論文では、この「視点分類タスク」に対し、近年盛ん

に研究が行われている単語埋め込みを用いた手法を提案する。特に、Twitter を対象とし、各投稿に付随するユーザーのプロフィール情報を用いることにより、分類精度の向上を目指す。類似する志向や立場を持つユーザーは、プロフィール情報に類似する単語を記述する傾向があることが期待できるため、単語埋め込みに、ユーザーの傾向を反映させ、より分類に有用なベクトル学習が可能になることが期待できる。

2. 関連研究

文章からのニューラル単語埋め込み獲得については、word2vec ([8]) や GloVe ([12]) を端緒とし、数多くの研究が行われている。通常、文章を対象とした単語分散表現の獲得においては、文章を単語のリストと捉え、各単語と周辺単語の関係を学習するモデルが採用されるが、本研究では、これをプロフィール情報を利用できる形に拡張する。

オンラインでの投稿のスタンスを分類するタスク (Stance Classification) は、主にオンラインディベートを対象に、様々な研究が提案されてきた。Somasmundaran らは、オンラインディベートのスタンス分類において、主観語辞書と文法的ルールにより対象と評価のペアを抽出し、「blackberry を好意的に評価する投稿はキーボードに言及する確率が高い」等を WWW からの Distant Supervision により学習する手法を提案した [17]。Hasan らは、議論構造と言語的素性に対し、生成・分類の両方のモデルを試している。[6] Sridhar らは、Probabilistic Soft Logic(PSL) を用いて、議論構造と言語的素性をモデル化することにより全体の構造を見ながら賛否分類を行っている。[18]。Anand らは、Unigram, Bigram, 反論か否か、依存構造、等の様々な素性を用いたスタンス分類を行った [1]。Ranade らは、sentiWordNet と、依存構造や談話マーカ (“but” 等) およびディベートの構造とを用いて、ユーザの意図や投稿の賛否分類を行っている [14]。Murakami らは、コメント間

¹ 徳島大学大学院社会産業理工学研究部
Tokushima University

a) mino@is.tokushima-u.ac.jp

b) matumoto@is.tokushima-u.ac.jp

c) kita@is.tokushima-u.ac.jp

のリンクに賛成や反対といったラベルを付与し、全体のリンク構造から各ユーザーのスタンスを抽出する手法を提案した [11]. Cohen らは、LDA や SVM を用いた既存研究を再検証し、一般性のある高精度な手法を構築することの難しさを報告している [2]. また、WWW 一般の文書を対象とした研究では、例えば、Mizuno らは、シソーラスや依存構造解析を用いて、クエリに対する WWW 上の文章の賛否を分類するシステムを提案している。 [9]

一方、近年では、SNS、特に Twitter を対象としたスタンス分類の研究が盛んとなっている。Rajadesingan らは、リツイート関係を利用したラベル伝搬による Twitter 投稿におけるスタンス分類を提案している [13]. Du らは、評価対象にとって重要な語を attention mechanism によって学習する手法を提案した [3]. Ebrahimi らは、Hinge-loss MRF を用いて、友人関係やツイートの類似性に基づく構造を反映しつつ、ツイートのラベルを学習する手法を提案した [5]. Sasaki らは、行列分解により「このテーマに賛成するユーザはこのテーマも賛成し易い」といった協調フィルタリングをスタンス分類において行う手法を提案している [16]. SemEval2016 では、ツイートのスタンス分類を行うタスクが提案されており、ラベル付きタスクでは、RNN を用いたシステムが、ラベル無しタスクでは、ルールベースのアノテーションと deep CNN を用いたシステムが、それぞれ最高精度を達成したと報告されている。 [10], Ebrahimi らは、SemEval2016 の Tweet スタンス分類に対し、Sentiment, Target, Stance を同時にモデル化する手法を提案している。 [4]

3. 問題設定

本研究では、入力を (t_i, p_i) のペアから構成されるリストとする。ここで、 t_i は各ツイート、 p_i はプロフィールの各文字列である。 t_i , p_i はそれぞれ単語に分解され、その各単語を t_{ij} , p_{ij} で表現する。各ペアを、ベクトルの学習に用いる。

本研究では、埋め込みの目的を、「プロフィール情報が与えられていないツイートに対し、より良いベクトルを与える」ものとする。このため、テストデータは、プロフィール情報を付随しないベクトル t_i のリストとし、各 t_i に対する埋め込みベクトルを出力とする。これにより、プロフィール情報が付随しないツイートを、視点の似たもの同士でまとめ、フィルタリングやサジェストに用いることが可能となる。この問題設定では、ツイートの意味ベクトルを、ツイートからの情報から構築する必要がある。このとき、ツイート中の単語ベクトルに、あらかじめ共起しやすいプロフィール単語の情報が含まれていれば、より良いベクトル表現を得ることができると考えられる。

4. 提案手法

提案手法では、word2vec と同様、Negative Sampling を利用した Skip-gram モデル (SGNS) を用いる。このため、本節では、まず SGNS について紹介し、その後その拡張として提案手法について述べる。

4.1 Skip-gram model with Negative Sampling (SGNS)

SGNS のスコアは、訓練データにおける、“input word” w_I からの “output word” w_O の予測確率をもとに定義される。word2vec の実装においては、 w_I が対象語、 w_O が文脈語となっている。各 (w_I, w_O) のペアに対するスコアは、以下で定義される。

$$\log \sigma(v'_{w_O} \cdot v_{w_I}) + \sum_{k=1}^K \log \sigma(-v'_{w_k} \cdot v_{w_I})$$

ここで、 K は Negative sample の数、 v_w と v'_w はそれぞれ、「入力ベクトル」「出力ベクトル」と呼ばれる二種類のベクトルであり、 σ はシグモイド関数である。Negative sample w_k は、各学習ステップにおいてランダムサンプリングされる。

SGNS では、この関数を最大化することにより分散表現ベクトル v_w と v'_w を得る。

4.2 プロフィール情報の単語ベクトルへの埋め込み

SGNS では「入力ベクトル」「出力ベクトル」という二種のベクトルを用いていたが、これに対し、提案手法では、「本文ベクトル」 v_w^s 、「本文文脈ベクトル」 v_w^{sc} 、「プロフィールベクトル」 v_w^p 、「プロフィール文脈ベクトル」 v_w^{pc} の四種のベクトルを用いる。

以下に、これら四種のベクトルについて説明する。なお、以下の説明では、 w を対象語、 z をその文脈語とする。

本文ベクトル・本文文脈ベクトル: 本文ベクトルと本文文脈ベクトルは、SGNS と同様のベクトルであり、「本文ベクトル」で表現された属性語を、「本文文脈ベクトル」で表現された周辺の語から推定する、という仮定に基づき、 v_w^s と v_w^{sc} の内積によりスコアを定義する。

プロフィールベクトル・プロフィール文脈ベクトル: 上記と同様、プロフィールベクトル v_w^p で表現された各語を、プロフィール文脈ベクトル v_w^{pc} で表現された文脈語から推定するというモデルに基づき、 v_w^p と v_w^{pc} の内積としてスコアを定義する。

SGNS と同様、各ツイートにおいて、対象語 w と文脈語 c のペアを一つずつとり、 w と c それぞれについて、四種のベクトルから一種が選択され、その間の重みを、スコアを増大させる方向に学習する。

4.3 目的関数

モデルの効果を測定するため、段階的に発展させた下記の四種のモデルを用いる。それぞれのモデルは、上で述べたベクトル同士の関係のうちいくつか（あるいは全て）を用いる。

4.3.1 本文単独モデル (モデル 1)

本文ベクトル v^a と本文文脈ベクトル v^{ac} 、及びその二ベクトル間の関係のみを考慮する。このため、目的関数は、SGNS の目的関数と同様の形となる。

本文の各語 w と、その文脈語 z それぞれに対し、スコア関数は

$$\log \sigma(v_z^{sc} \cdot v_w^s) + \sum_{k=1}^K \log \sigma(-v_z^{sc} \cdot v_{w_k}^s)$$

と定義される。ここで、 K は Negative sample の数、 w_k は一様分布から選ばれた疑似負例である。一つのペアに対する目的関数は、本文中の各 w_i について、同一のツイート中の各単語を z とし、上記を全 (w, z) のペアについて足し合わせた、以下の式となる。

$$l = \sum_{w \in t_i} \sum_{z \in t_i, z \neq w} \{ \log \sigma(v_z^{sc} \cdot v_w^s) + \sum_{k=1}^K \log \sigma(-v_z^{sc} \cdot v_{w_k}^s) \}$$

4.3.2 本文-プロフィール関連モデル (モデル 2)

本文-プロフィール関連モデルでは、本文とプロフィールそれぞれの中で単語と文脈の関係を学習するほか、本文中単語とプロフィール中単語の間についても学習する。これを、本文文脈ベクトルを、本文ベクトルとプロフィールベクトル両方の推定に用いることで実現する。本文の各語 w に対して、その文脈語 z を本文から取得する部分はモデル 1 と同様であるが、プロフィール中の各語に対して、その文脈語を、同じプロフィール中から取得するほか、本文からも取得する。すなわち、プロフィール中の各語を、同一プロフィール中の語と、付随する本文中の語の二種の文脈から予測するモデルとなっている。プロフィール中の各語 w と、その文脈語 z に対し、スコア関数は、

$$l_{(z,w)} = \begin{cases} \log \sigma(v_z^{sc} \cdot v_w^p) + \sum_{k=1}^K \log \sigma(-v_z^{sc} \cdot v_{w_k}^p) & (z \text{ が本文語の場合}) \\ \log \sigma(v_z^{pc} \cdot v_w^p) + \sum_{k=1}^K \log \sigma(-v_z^{pc} \cdot v_{w_k}^p) & (z \text{ がプロフィール語の場合}). \end{cases}$$

と定義され、全体の目的関数は、これらを全ての (w, t) について足し合わせた、以下の形となる。^{*1}

^{*1} プロフィールベクトルは、目的関数中で二種のベクトル（本文文脈ベクトル及びプロフィール文脈ベクトル）と関係しているが、学習は word2vec と同様に、各ステップで 1 つの各単語・文脈ペア (w, z) のみを訓練データと見なした SGD により行っており、二項のうち一項のみが有効となるため、word2vec と同様の更新式が使えることに注意されたい。

$$l = \sum_{w \in p_i} [\sum_{z \in p_i, z \neq w} \{ \log \sigma(v_z^{pc} \cdot v_w^p) + \sum_{k=1}^K \log \sigma(-v_z^{pc} \cdot v_{w_k}^p) \} \\ + [\sum_{z \in t_i} \{ \log \sigma(v_z^{sc} \cdot v_w^p) + \sum_{k=1}^K \log \sigma(-v_z^{sc} \cdot v_{w_k}^p) \}] \\ + \sum_{w \in t_i} [\sum_{z \in t_i, z \neq w} \{ \log \sigma(v_z^{sc} \cdot v_w^s) + \sum_{k=1}^K \log \sigma(-v_z^{sc} \cdot v_{w_k}^s) \}]]$$

4.3.3 重み付けモデル

さらに、上記を拡張するモデルとして、ツイート中の各単語に、予測に有用な単語とそうでない単語があるという仮定のもと、本文文脈ベクトル $v^{sc}(w)$ を利用する際、それに付随する重み $a(w)$ を仮定し、内積にその値を掛けるというモデルを提案する。これにより、本文文脈単語のうち、重要な単語に着目して予測を行うというアテンションの効果が期待できる。このとき、目的関数は、モデル 1、モデル 2 と同、 v_z^{sc} の登場する内積にさらに a_z を乗じたものとなる。

4.4 パラメータ学習

パラメータの学習は、word2vec と同様、一つの単語・文脈ペア (w, c) 毎に更新を行う確率的勾配降下法 (SGD) を用いる ([7])。また、やはり word2vec と同様、並列化に、データセットを分割して単純に各プロセッサを割り当て、競合を無視する Hogwild 法 ([15]) を用いた。学習の反復数は 10 とした。

5. 実験

Twitter API を利用し、2017 年 7 月から 10 月にかけてランダムサンプリングにより収集したデータを使用する。このうち、リツイートの記号 (“RT @”) が含まれているか、URL 引用 (“http...”) が含まれているもののみを解析の対象とする。

頻度 10 以上の単語についてベクトルを学習した。上記四種のモデルの比較のほか、比較対象として、単語をそのまま次元として用いる通常のベクトル空間モデルも用いる。また、次元数は 50、Negative Sample の数 K は 25 とした。

本論文では、制度計測のため、クラスタリングを用いたツイートラベリングタスクを提案する。テストデータは、テストセット中のツイートをを用い、付随するプロフィール情報は用いない。ツイート中の各単語を平均（重み付きモデルでは重み付き平均）することにより、テストデータ中の各ツイートを表すベクトルを得る。得られたベクトルをもとに、K-means 法を用いてベクトルをクラスタリングすることで、各ベクトルに、クラスタの ID をラベリングすることができる。

ツイートラベリングタスクは、「同一ユーザーは同一の視点を保持する」という仮定のもと、複数のツイートを投稿しているユーザーに着目し、それらのツイートに同一のラベルを付与できるかどうかの精度を計測する。

あるユーザーの投稿のリストを T_u とし、 T_u 中の各ツイートにラベル付を行う。付与されたラベルのうち、最も多かったものを l_{max} とし、このとき、 T_u へのラベリングの純度 (purity) を、

$$\frac{l_{max}}{|T_u|}$$

と定義する。(ただし、 $|T_u|$ は T_u の要素数)。この値をテストデータ中のすべてのユーザーに対し平均したものを、平均純度 (average purity) として評価に用いる。

本実験では、「保育園」、「徳島」の2つのクエリで検索して得られたツイート集合を実験対象とする。「保育園」は、総ツイート数 7493 ツイート、学習用ツイート 3195 種、テストデータ 140 ユーザであり、「徳島」は、総ツイート数 8401 ツイート、学習用ツイート 4648 種、テストデータ 290 ユーザであった。それぞれのツイート集合に対し、次に述べるツイートラベリングタスクを行い、各手法の精度を比較する。計測は、各手法について3回行い、その平均を計算する。

5.1 結果

表1および2に結果を示す。今回のクエリに対しては、プロフィール情報を用いた提案手法が、概ね高い精度を示した。また、重み付けを用いた場合は、特にクエリ「徳島」の場合に、精度の向上が見られた。これにより、プロフィール情報の利用および重み学習に、一定の効果があることが確認できた。

6. おわりに

同一の単語に関するツイートを視点により分類する「視点分類タスク」を提案し、そのための手法として、プロフィール情報を援用した単語埋め込み法を提案した。2種類のクエリを用いた実験では、プロフィール情報の利用と、単語重み学習に一定の効果があることが確認できた。今後は、より様々なクエリに対する、各手法の効果の確認を行なっていく予定である。

謝辞 本研究は JSPS 科研費 15K00425, 15K00309, 15K16077 の助成を受けたものです。

参考文献

- [1] Anand, P., Walker, M., Abbott, R., Tree, J. E. F., Bowman, R. and Minor, M.: Cats Rule and Dogs Drool!: Classifying Stance in Online Debate, *Proceedings of 2nd Workshop on Computational Approaches to Subjectivity and Sentiment Analysis (WASSA 2.011)*, pp. 1–9 (2011).
- [2] Cohen, R. and Ruths, D.: Classifying Political Orienta-

- tion on Twitter: It's Not Easy!, *Proceedings of ICWSM 2013*, pp. 91–99 (2014).
- [3] Du, J., Xu, R., He, Y. and Gui, L.: Stance Classification with Target-specific Neural Attention, *Proceedings of IJCAI 2017*, pp. 3988–3994 (2017).
- [4] Ebrahimi, J., Dou, D. and Lowd, D.: A Joint Sentiment-Target-Stance Model for Stance Classification in Tweets, *Proceedings of COLING 2016*, pp. 2656–2665 (2016).
- [5] Ebrahimi, J., Dou, D. and Lowd, D.: Weakly Supervised Tweet Stance Classification by Relational Bootstrapping, *Proceedings of EMNLP 2016*, pp. 1012–1017 (2016).
- [6] Hasan, K. S. and Ng, V.: Stance Classification of Ideological Debates: Data, Models, Features, and Constraints, *Proceedings of IJCNLP 2013*, pp. 1348–1356 (2013).
- [7] Ji, S., Satish, N., Li, S. and Dubey, P.: Parallelizing Word2Vec in Shared and Distributed Memory, *CoRR abs/*, Vol. 1604.04661 (2016).
- [8] Mikolov, T., Sutskever, I., Chen, K., Corrado, G. S. and Dean, J.: Distributed Representations of Words and Phrases and their Compositionality, *Proceedings of NIPS 2013*, pp. 3111–3119 (2013).
- [9] Mizuno, J., Nichols, E., Watanabe, Y. and Inui, K.: Organizing Information on the Web through Agreement-Conflict Relation Classification, *Proceedings of AIRS 2012*, pp. 126–137 (2012).
- [10] Mohammad, S., Kiritchenko, S., Sobhani, P., Zhu, X.-D. and Cherry, C.: SemEval-2016 Task 6: Detecting Stance in Tweets, *Proceedings of SemEval@NAACL-HLT 2016*, pp. 31–41 (2016).
- [11] Murakami, A. and Raymond, R.: Support or Oppose? Classifying Positions in Online Debates from Reply Activities and Opinion Expressions, *Proceedings of COLING 2010 (Posters)*, pp. 869–875 (2010).
- [12] Pennington, J., Socher, R. and Manning, C. D.: Glove: Global Vectors for Word Representation, *Proceedings of EMNLP 2014*, pp. 1532–1543 (2014).
- [13] Rajadesingan, A. and Liu, H.: Identifying Users with Opposing Opinions in Twitter Debates, *Proceedings of SBP 2014*, pp. 153–160 (2014).
- [14] Ranade, S., Sangal, R. and Mamidi, R.: Stance Classification in Online Debates by Recognizing Users' Intentions, *Proceedings of SIGDIAL Conference 2013*, pp. 61–69 (2013).
- [15] Recht, B., Re, C., Wright, S. J. and Niu, F.: Hogwild: A Lock-Free Approach to Parallelizing Stochastic Gradient Descent, *Proceedings of NIPS 2011*, pp. 693–701 (2011).
- [16] Sasaki, A., Hanawa, K., Okazaki, N. and Inui, K.: Other Topics You May Also Agree or Disagree: Modeling Inter-Topic Preferences using Tweets and Matrix Factorization, *Proceedings of ACL 2017 (1)*, pp. 398–408 (2017).
- [17] Somasundaran, S. and Wiebe, J.: Recognizing Stances in Online Debates, *Proceedings of ACL/IJCNLP 2009*, pp. 226–234 (2009).
- [18] Sridhar, D., Getoor, L. and Walker, M.: Collective Stance Classification of Posts in Online Debate Forums, *Proceedings of ACL Joint Workshop on Social Dynamics and Personal Attributes in Social Media*, pp. 109–117 (2014).

表 1 クエリ=“保育園”に対する結果. “提案手法”はモデル 2 を, “本文のみ”はモデル 1 を示す. (W) は重み付きモデルを示す.

クラス数	提案手法 (W)	提案手法	本文のみ (W)	本文のみ	ベースライン
2	91.21	92.83	90.78	90.9	77.82
3	84	87.25	86.48	83.57	74.72
4	81.95	79.67	78.68	78.39	74.2
5	77.57	79.32	76.68	77.6	70.72
6	75.68	77.59	74.86	77.42	69.84
7	74.02	73.29	73.88	73.96	68.6
8	72.08	74.89	73.5	71.99	69.16
9	74.58	73.22	72.02	70.48	66.81
10	70.78	69.9	71.1	69.89	67.11

表 2 クエリ=“徳島”に対する結果. “提案手法”はモデル 2 を, “本文のみ”はモデル 1 を示す. (W) は重み付きモデルを示す.

クラス数	提案手法 (W)	提案手法	本文のみ (W)	本文のみ	ベースライン
2	98.61	93.36	90.81	87.54	87.74
3	93.94	85.83	87.87	86.18	82.57
4	87.51	83.42	86.68	81.2	82.53
5	85.59	82.71	84.33	80.92	80.4
6	83.62	82.72	84.24	77.95	77.55
7	81.76	78.7	81.02	77.52	76.91
8	80.88	77.16	79.14	76.63	75.56
9	77.28	77.02	77.46	75.13	76.02
10	79.27	75.03	75.82	75.15	73.44