

半教師あり学習を用いたテキストデータ拡張による フェイクニュース検出

渡辺 春希[†] 森 康真[†] 田村 慶一[†]

[†] 広島市立大学大学院情報科学研究科 〒731-3194 広島県広島市安佐南区大塚東 3-4-1

E-mail: [†] {mh67026@e., mori@, ktamura@}hiroshima-cu.ac.jp

あらまし ソーシャルメディアの普及によりフェイクニュースと呼ばれる事実と異なる情報が広がり、多くの人を混乱に巻き込んでいる。このようなフェイクニュースの拡散に対し、機械学習や自然言語処理の技術を用いて自動でテキストの真偽を分類するフェイクニュース検出と呼ばれる研究が行われている。フェイクニュース検出の課題点として分類したい話題によっては十分なデータが存在しないことが挙げられる。そのため少ないデータ量でもモデルを十分に学習させる手法が必要である。本研究ではテキストデータの拡張を行いデータ量の課題点を解決させ、モデルに十分な学習を行わせる。更に拡張されたテキストデータに対して半教師あり学習を用いてラベル付与を行い、精度向上に影響があるか調査を行う。

キーワード 自然言語処理, 機械学習, 半教師あり学習, フェイクニュース検出, テキスト分類, ニュース・マスコミ

1. はじめに

SNSの普及により多くの人々が手軽に情報を発信したり、様々な情報を取得することが可能となった。その一方で信憑性の欠けた情報であるフェイクニュースも拡散されるようになった。フェイクニュースとして扱われる事象は、デマや陰謀論、風刺など様々な定義がある[1]。フェイクニュースの目的は衝撃的なタイトルで人々を欺き閲覧数を稼ぎ広告収入を得たり、特定の人物や団体に対して誹謗中傷をすることである。2016年のアメリカ大統領選挙では信頼性の低い情報がTwitterで拡散され、選挙の結果に影響を与えた可能性がある¹とされる[2]。日本においてもフェイクニュースの被害として東日本大震災に関する出来事[3]やトイレットペーパー不足などが挙げられる。

そのようなフェイクニュースの被害を最小限に抑えるための対抗策としてファクトチェックサイトの事実検証が挙げられる。ファクトチェックサイトの例として海外ではPolitifact¹やSnopes²が存在し、日本では日本ファクトチェックセンター³(JFC)などが存在する。しかし手動によるファクトチェックは専門知識と多大な労力を要するため、事実検証にかなりの時間が掛かりリアルタイムで処理することが難しいことが懸念点として挙げられる。そのため現在自然言語処理や機械学習の技術を用いてフェイクニュースを自動で分類する研究、フェイクニュース検出が行われている。

フェイクニュース検出の課題点として分類を行う

話題によってはデータ量が少ないことが挙げられる。マイナーな話題であれば対象の話題について取り上げている人が少ないため、多量のデータを取得することができない。また初期の新型コロナウイルスのように多くの人が取り上げていても得体の知れない未知の情報のため、事実確認された十分な情報が存在しない。そのためこのようなフェイクニュース検出の課題点を解決するために、少ない学習データでフェイクニュース検出モデルを学習させる手法や技術が必要である。少ない学習データでモデルを学習させる手法としてデータの拡張、転移学習、Few-shot Learningなどが挙げられる。

本研究では既存のテキストデータの拡張に改良を行いフェイクニュース検出モデルに十分な学習を行わせる。さらに、拡張されたテキストのラベルをオリジナルのテキストを基に割り当てた場合と、半教師あり学習を用いてラベルを割り当てた場合ではどちらの方が精度が優れているかについて実験を行う。

2. 関連研究

本研究ではテキストデータの拡張手法であるEDA(Easy Data Augmentation)[4]、PDA(Part of speech Data Augmentation)[5]と自然言語処理分野の半教師あり学習手法であるDelta-training[6]を用いる。2.1節と2.2節ではEDAとPDAの紹介を行い、2.3節ではDelta-trainingの紹介を行う。

¹ Politifact : <https://www.politifact.com/>

² Snopes : <https://www.snopes.com/>

³ JFC : <https://factcheckcenter.jp/>

2.1 EDA

EDA はテキスト中の単語に対し変換を行い、テキストデータの拡張を行う。EDA の概要を図 1 に示す。

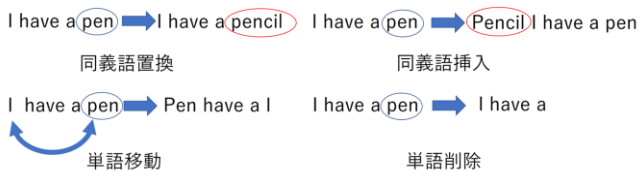


図 1 :EDA の概要

EDA による具体的な単語の変換方法として以下の 4 つが存在する。

- ・同義語置換：テキスト中の単語を同義語に置き換える。
- ・同義語挿入：テキスト中の単語を選択し、その単語の同義語をランダムな位置に挿入する。
- ・単語移動：テキスト中の単語を選び位置を交換する。
- ・単語削除：一定の確率でテキスト中の単語を削除する。

同義語置換、同義語挿入、単語移動に関しては処理を N 回行う。 $N = \alpha l$ を満たし、 α は単語の変換率、 l はテキスト中の単語数を表す。単語削除において α は単語を削除する確率を表す。

2.2 PDA

EDA では変換を行った単語の品詞によってはテキストの意味が大きく変化し、拡張されたテキストに対応するラベルが一致しなくなる可能性がある。例えば変換された単語の品詞が名詞の場合、テキストの主語や目的語が変化する。PDA では予め指定された品詞の単語について変換を行わない。ここでは同義語置換を例に PDA の概要を図 2 に示す。

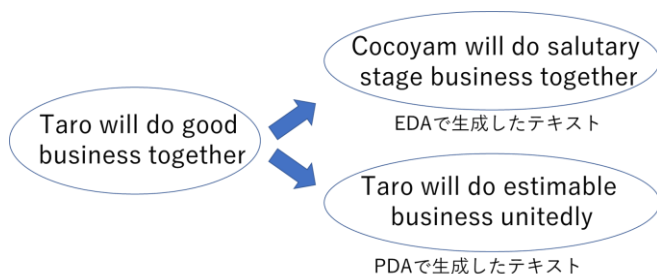


図 2:PDA の概要

変換を行わない単語に名詞を指定した場合 Taro と business には変換が行われず、他の単語が変換される。

2.3 Delta-training

Delta-training は 2 つのモデルの性能差を利用した半教師あり学習である。2 つのモデルは同じモデルを使い、単語の分散表現を変化させて性能差を測る。優れたモデルは Glove[7]の事前学習済み分散表現を用い、劣ったモデルはランダムに初期化された分散表現を用いる。Delta-training の概要を図 3 に示す。

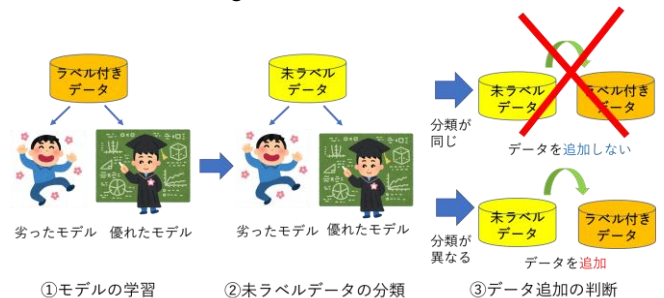


図 3:Delta-training の概要

Delta-training の流れは以下の通りである。

- ① 優れたモデルと劣ったモデルを用意し、ラベル付きデータを用いてモデルの学習を行う。
- ② 学習させたそれぞれのモデルを用いて未ラベルデータの分類を行う。
- ③ 対象の未ラベルデータの分類結果がそれぞれのモデルで異なった場合、優れたモデルの分類結果を採用し対象の未ラベルデータをラベル付きデータに追加する。対象の未ラベルデータの分類結果がそれぞれのモデルで同じならば、対象の未ラベルデータをラベル付きデータに追加しない。

Delta-training は①~③のステップを 5 回繰り返し、未ラベルデータをラベル付きデータに追加する。

3. 提案手法

本研究の提案手法を図 4 に示す。

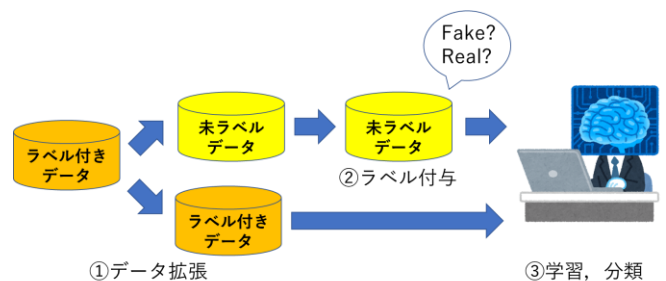


図 4: 提案手法の流れ

提案手法の流れは以下の通りである。

- ① EDA もしくは PDA によりラベル付きデータを用いてテキストデータの拡張を行う。拡張されたテキストデータに対してはラベルの付与は行わな

い.

- ② Delta-training を用いて拡張された未ラベルデータに対して、ラベルの付与を行う.
- ③ 元々のテキストデータと拡張されたテキストデータを用いてモデルに十分な学習を行わせ、テストデータに対して分類を行う.

4. 評価実験

本章では提案手法による分類精度の評価実験を行う.

4.1 使用データ

使用データはシリア戦争のニュース記事についてまとめた FA-KES[8]を使用する. FA-KES は 426 件のリアルニュースと 378 件のフェイクニュースから構成されている. データセットのラベル付けに関しては主観性を排除し, 正確に判断するために人手で行わず, 半教師あり学習を用いてデータセットのラベル付けを行っている.

4.2 実験設定

FA-KES データセットの記事の日付順にソートを行い, 全体の 80%である古い記事を拡張のための訓練データとし, 残りの 20%の新しい記事をテストデータとする. またテキストデータ拡張の EDA, PDA のパラメータは以下の通りである.

- データを増やした数: 2 倍, 3 倍, 4 倍, 5 倍
- 品詞の指定: 名詞(NN), 動詞(VB)
- 操作: 全て(ALL), 単語削除(RD), 同義語置換(SR), 単語削除と同義語置換(RD&SR)
- α : 0.05, 0.10, 0.50

使用モデルには CNN と Bi-LSTM を用いる.

4.3 実験方法

拡張されたテキストデータに対して, オリジナルのテキストを基にラベル付けを行った場合と Delta-training を基にラベル付けを行った場合の分類精度の比較を行う. オリジナルのテキストを基にラベル付けを行う場合は, オリジナルのテキストのラベルが Fake ならば拡張されたテキストのラベルを Fake とする. 分類精度の評価指標には正解率, 適合率, 再現率, F 値を用いる.

4.4 実験結果

EDA でオリジナルのテキストを基にラベル付けを行った場合の正解率上位 5 件の結果を表 1 に示す. パラメータの列の値は左から順に行った操作, α の値, データを増やした数, 使用モデルである.

表 1 : Delta-training を用いなかった場合の EDA の上位 5 件

パラメータ	正解率	適合率	再現率	F 値
RD&SR,0.50,5 倍,Bi-LSTM	57.0	55.2	54.1	52.2
RD&SR,0.50,4 倍,Bi-LSTM	55.9	54.6	54.1	53.6
SR,0.50,5 倍,Bi-LSTM	55.8	55.0	54.3	53.5
SR,0.50,2 倍,Bi-LSTM	55.8	54.5	53.3	51.5
SR,0.10,2 倍,Bi-LSTM	55.5	53.7	52.6	50.6

EDA で Delta-training を基にラベル付けを行った場合の正解率上位 5 件の結果を表 2 に示す. パラメータの列の値は左から順に行った操作, α の値, データを増やした数, 使用モデルである.

表 2 : Delta-training を用いた場合の EDA の上位 5 件

パラメータ	正解率	適合率	再現率	F 値
RD&SR,0.50,4 倍,Bi-LSTM	59.0	58.2	56.5	55.2
RD&SR,0.10,5 倍,Bi-LSTM	58.9	58.0	56.7	55.8
SR,0.10,3 倍,CNN	56.6	58.4	51.8	43.1
ALL,0.05,4 倍,Bi-LSTM	56.5	55.8	53.6	51.2
RD&SR,0.50,4 倍,Bi-LSTM	56.5	54.4	52.2	46.2

それぞれの最高正解率を用いて比較をした場合, Delta-training を用いることで最高正解率が 2.0%向上した.

次に PDA でオリジナルのテキストを基にラベル付けを行った場合の正解率上位 5 件の結果を表 3 に示す. パラメータの列の値は左から順に行った操作, α の値, 指定した品詞, データを増やした数, 使用モデルである.

表 3 : Delta-training を用いなかった場合の PDA の上位 5 件

パラメータ	正解率	適合率	再現率	F 値
SR,0.10,NN,3 倍,Bi-LSTM	56.8	55.7	55.2	54.6
SR,0.05,VB,4 倍,Bi-LSTM	56.5	55.5	54.9	54.6
SR,0.05,VB,2 倍,Bi-LSTM	55.8	54.0	53.3	52.1
SR,0.10,VB,4 倍,Bi-LSTM	55.8	54.6	54.4	54.2
ALL,0.50,VB,3 倍,CNN	55.5	53.6	52.8	50.8

PDA で Delta-training を基にラベル付けを行った場合の正解率上位 5 件の結果を表 4 に示す。パラメータの列の値は左から順に行った操作、 α の値、指定した品詞、データを増やした数、使用モデルである。

表 4 : Delta-training を用いた場合の PDA の上位 5 件

パラメータ	正解率	適合率	再現率	F 値
RD,0.50,NN,4 倍,Bi-LSTM	58.1	57.9	57.7	57.5
RD,0.50,VB,4 倍,Bi-LSTM	57.5	56.2	53.3	48.2
SR,0.10,NN ,3 倍,CNN	57.3	64.4	52.4	42.9
RD&SR,0.10,VB ,4 倍,Bi-LSTM	57.0	55.4	53.9	51.8
RD,0.10,NN,3 倍,CNN	56.4	52.3	52.2	46.0

それぞれの最高正解率を用いて比較をした場合、Delta-training を用いることで最高正解率が 1.3%向上した。

EDA と PDA での Delta-training の結果を比較すると、EDA の方が最高正解率が高く、正解率の上昇も EDA の方が高かった。これはデータ拡張の性質が原因だと考えられる。EDA は変換する単語に制限をかけていないため PDA と比較すると、様々なテキストが生成されやすく、拡張されたテキストの意味はオリジナルのテキストと異なりやすくなる。そのため拡張されたテキストのラベルをオリジナルのテキストと同じにすると、モデルが適切な学習を行えない可能性がある。よって Delta-training を用いて何らかの根拠に基づいたラベル付けを行うことで、モデルが幅広いテキストデータを適切に学習でき精度向上に繋がったと考えられる。

5. おわりに

本研究では FA-KES データセットを対象にテキストデータの拡張を行い、拡張されたテキストデータに対して半教師あり学習を用いてラベル付けを行う、フェイクニュース検出のためのテキストデータ拡張手法を提案した。テキストデータの拡張には EDA, PDA を用いて、半教師あり学習には Delta-training を用いた。Delta-training を用いてラベル付けを行った場合は PDA の方で最高正解率が 1.3%向上し、EDA の方では最高正解率が 2.0%向上した。

今後の課題としては他のフェイクニュースのデータセットでも精度向上が見られるか実験を行うことである。FA-KES データセットでは 1 件当たりの単語数が多かったが、Twitter のように 1 件当たりの単語数が少ないデータセットでは結果が変わる可能性がある。

また PDA, EDA 以外のテキストデータ拡張がフェイクニュース検出において有効かどうかについても調べる必要がある。

参 考 文 献

- [1] X. Zhou and R. Zafarani. A Survey of Fake News: Fundamental Theories, Detection Methods, and Opportunities. *ACM Computing Surveys*, vol. 53, no. 5, pp. 1-40, 2020.
- [2] Alexandre Bovet and Hern´an A Makse. Influence of fake news in twitter during the 2016 us presidential election. *Nature communications* 10(1), pp. 1-14, 2019.
- [3] Misako Takayasu et al. Rumor diffusion and convergence during the 3.11 earthquake: a twitter case study. *PLoS one* 10(4), pp. 1-18, 2015.
- [4] Jason Wei and Kai Zou. EDA : Easy Data Augmentation techniques for boosting Performance on text classification tasks. In *Proc. EMNLP-IJCNLP’19*, pp. 6382-6388, 2019.
- [5] 渡辺春希, 森康真, 田村慶一. 単語の品詞情報を用いたデータ拡張によるフェイクニュース検出の精度評価. 第 26 回日本知能情報ファジイ学会中国・四国支部大会講演論文集, pp. 46-47, 2022.
- [6] Hwiyeol Jo and Ceyda Cınarel. Delta-training: Simple Semi-Supervised Text Classification using Pretrained Word Embeddings. In *Proc. EMNLP-IJCNLP’ 19*, pp. 3458-3463, 2019.
- [7] Jeffrey Pennington et al. Glove: Global vectors for word representation. In *Proc. EMNLP’ 14*, pp. 1532-1543, 2014.
- [8] Fatima K Abu Salem et al. FA-KES: A fake news dataset around the Syrian war. In *Proc. International AAAI Conference on Web and Social Media*, Vol.13, pp. 573-582, 2019.