

深層感情分析に基づくデマ検知手法の開発 Rumor Detection based on Deep Sentiment Analysis

QU LIJING[†] 成 凱[‡]
Lijing QU Kai CHENG

1. はじめに

SNS では誰でもリアルタイムで受発信することができるため、情報源の一つとして利用されている。一方、自由に発信された情報の中に正しい情報だけでなく、事実誤認や根拠のないウワサ話、流言飛語といったデマも含まれており、社会問題となって対策が必要である。デマは様々な種類が存在するが、本研究では SNS 上で一般ユーザによって投稿され、社会問題や話題に対する他人及び組織に悪意を持って拡散される情報を対象とする。

デマ対策の一環として、デマ検知を素早く行う必要がある。近年、深層学習と感情分析がデマ検知に使われている。ディープニューラルネットワークにおける表現学習は従来の特徴量エンジニアリングよりよい特徴表現を抽出することができる。感情分析はテキスト解析とスタンス分類が重要である。デマは目新しく感じられ、接した人が驚きや恐れ、嫌悪感などを抱く特徴があるため、感情分析によってデマを効果的に検知できると期待される。

しかし、テキスト解析による感情分析は精度が低く、それに基づくデマ検知では誤検出の割合が高いという問題が指摘されている。また、スタンス分類がデマ検知に有効であるが、ユーザ信頼性分析を考慮したスタンス分類がデマ検知への有効性はまだ検証されていない。

本研究は既存手法を検証し、それぞれの長所と短所を明らかにする。深層学習に基づく感情分析(深層感情分析)を用いてデマの特徴量を抽出し、分類モデルを構築し、デマ検知の精度を高める。

2. デマ検知の主な技術

デマ対策のプロセスはデマ検知、デマ追跡、スタンス分類とデマ検証という一連のタスクから構成される^[3]。その中、デマ検知の既存手法は単一タスク学習に基づく方法とマルチタスク学習に基づく方法がある。

単一タスク学習に基づく方法はデマ検知を機械学習の二項分類タスクと見なす。分類性能はデータの表現に依存するため、特徴量の抽出は重要な部分となり、分類精度に影響を与える。デマ検知でよく使われている特徴量はコンテンツ情報、ユーザ情報と拡散メカニズムとネットワークに基づく特徴量である。Vosoughi ら(2015)^[1]はテキストコンテンツの言語スタイル、ユーザの特性とネットワーク伝播の時系列特徴を用いてデマ真実性を予測した。Azri ら(2021)^[2]は画像コンテンツと感情表現を注目し、deepMONITORを開発した。

マルチタスク学習に基づく方法は共有表現を持つメインタスクと補助タスクの共同学習と指す。複数の異なるタスク間の共通の要因を獲得し、互いにヒントを与える。

[†]九州産業大学大学院情報科学研究科 Graduate School of Information Science, Kyushu Sangyo University

[‡]九州産業大学理工学部 Faculty of Science and Engineering, Kyushu Sangyo University

Kochkina ら(2018)^[3]はデマ検証をメインタスクとし、デマ検出、スタンス分類を補助タスクとしてデマ検証のパフォーマンスを向上させた。CHEN ら(2021)^[4]はスタンス分類タスクと早期デマ検出タスクの共同学習を行った。

3. 提案手法

提案手法は図 1 で示す。データを前処理し、深層感情分析によるテキスト特徴量、感情特徴量とユーザ信頼性を考慮したスタンス特徴量を抽出し、分類モデルで検知する。

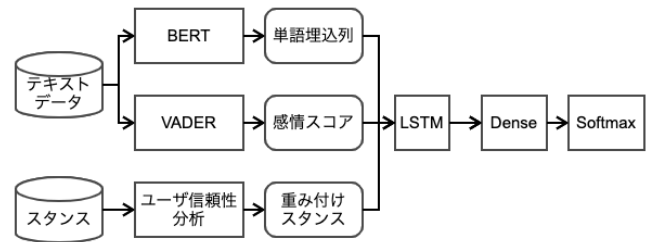


図 1 提案手法の流れ

3.1 データの前処理

アルファベット以外の文字、ストップワード辞書より不要な句読点、スペースを削除し、テキストを正規化する。

3.2 深層感情分析による特徴量抽出

(1) テキスト特徴量

テキスト特徴量はデマ検知によく使われている。事前学習済みモデル BRRT を用いて単語埋め込み表現を抽出する。

(2) 感情特徴量

VADER を用いて感嘆符、顔文字を考慮し、辞書とルールの組み合わせで感情極性と強度を抽出する。

(3) ユーザ信頼性を考慮したスタンス特徴量

信頼性の高いユーザと信頼性の低いユーザのスタンス両方とも重要である。「信頼できないユーザの支持」と「信頼できるユーザの否定、疑い」は統計的なスタンス分析(支持、否定の数など)よりもデマの真実性を反映できる。各スタンス(SDQC)の信頼性を分析し、重み付ける。

4. 予備実験

予備実験ではスタンス分類がデマ検知への有効性の検証実験を行った。具体的に、マルチタスク学習を用いてデマの真実性分類(メインタスク)とスタンス分類(補助タスク)実験を行った。最後、多数決(Majority Vote)を用いたデマの真実性分類結果と比較した。

4.1 概要

まず、モデルのパラメータ最適化を行った。具体的に、Tree-structured Parzen Estimator Approach(TPE), (NOPS,2011)手法を使用し、hyperoptで実装した。hyperoptはパラメータ探索を効率よく行う python ライブラリであり、探索するパ

ラメータ空間を事前に設定する必要がある。パラメータの設定と最適化した結果は表 1 のように示す。

表 1 パラメータ最適化

パラメータ	設定	結果
バッチサイズ	32	32
L2 正則率	{0.001,0.001}	0.001
学習率	0.0001	0.0001
Dense 層の数	{1,2}	2
Dense 層のユニットの数	{300, 400, 500, 600}	500
訓練回数	50	50
LSTM 層の数	{1,2}	1
LSTM 層のユニットの数	{100, 200, 300}	300

次に、最適化したパラメータを採用し、keras と sklearn を使用し、図 2 のように、Dense 層は 2 層、ユニット数 500、LSTM 層は 1 層、ユニット数 300 のニューラルネットワークを構築した。また、スタンス分類はツイッターレベルなので、共有層 LSTM 以外に LSTM 層を追加しない。

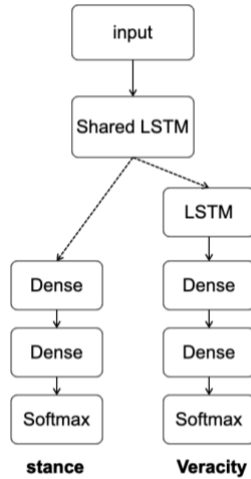


図 2 マルチタスクのモデル

4.2 データセット

SemEval-2019 Task7 : Determining Rumor Veracity and Support for Rumors で提供されたデータセットを使用する。

このデータセットは 9 種類異なる話題に関連する Twitter の会話スレッド (Conversation threads) が含まれている。各会話スレッドは真、偽または未検証のラベル付きのソース投稿 (source) とスタンスのラベル付きの返信 (reply) から構成される。一つの会話は幾つのブランチ (branches) がある。

図 3 では「ottawa shooting」という話題のデマ会話例の一つを示す。

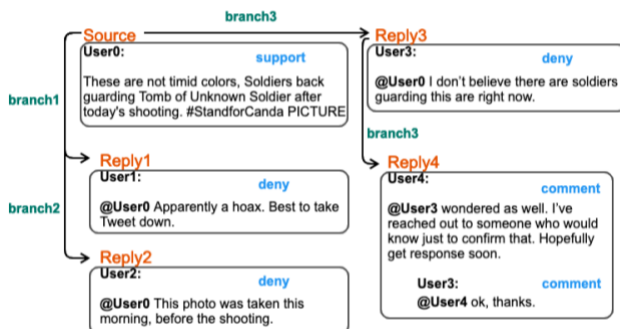


図 3 会話スレッド例

4.3 評価方法

正解率 (accuracy), 適合率 (precision), 再現率 (recall) と Marco F1 値を用いて分類モデルの性能を評価する。

正解率 (a) は全予測正答率であり、以下のように示す。

$$a = \frac{\text{No. of rumors and norumors predicted correctly}}{\text{No. of rumors and norumors}} \quad (1)$$

クラス C に対して、適合率 (p), 再現率 (r) と $F1$ は以下のように示す。

$$p = \frac{\text{No. of rumors predicted as } C \text{ correctly}}{\text{No. of rumors predicted as } C} \quad (2)$$

$$r = \frac{\text{No. of rumors predicted as } C \text{ correctly}}{\text{No. of rumors annotated as } C} \quad (3)$$

$$F1 = \frac{2 * p * r}{p + r} \quad (4)$$

一部のデータセットは偏っているため、Macro F1 はクラス全体の性能を表すことができる。

$$\text{Macro F1} = \frac{1}{n} \sum_{i=1}^n F1_i \quad (5)$$

なお、 n はクラスの数、 $F1_i$ は各クラスの $F1$ 。

4.4 実験結果

データセットの「Ferguson」, 「Germanwings Crash」, 「Ottawa Shooting」, 「Sydney Siege」に関する話題を訓練データとし、「Charlie Hebdo」に関する話題をテストデータとして多数決とマルチタスク学習を用いたモデルの分類結果を表 2 にまとめた。Task A はスタンス分類、Task B はデマ真実性分類と指す。

表 2 結果

	accuracy	precision	recall	Macro F1
Majority	0.286	0.279	0.103	0.148
Task A	0.626	0.368	0.374	0.365
Task B	0.407	0.405	0.435	0.398

表 2 の正解率と Macro F1 を見ると、Majority より、マルチタスクを用いたデマ真実性分類 (Task B) の性能を向上させることができると検証した。

5. 終わりに

本研究ではデマ検知の既存手法を調査した。深層感情分析による特徴量を抽出し、分類モデルでデマを検知するという手法を提案した。予備実験ではマルチタスク学習を用いてスタンス分類とデマ真実性分類の実験を行い、デマ検知への有効性を検証した。今後の予定は学習モデルを構築し、デマ検知の精度を高める。

参考文献

- [1] Soroush Vosoughi. "Automatic Detection and Verification of Rumors on Twitter". Ph.D. Dissertation (2015)
- [2] Azri, Abderrazek, et al, "Calling to CNN-LSTM for Rumor Detection: A Deep Multi-channel Model for Message Veracity Classification in Microblogs", Joint European Conference on Machine Learning and Knowledge Discovery in Databases. Springer, Cham, (2021)
- [3] Kochkina, E., Liakata, M., & Zubiaga, A., "All-in-one: Multi-task learning for rumour verification", arXiv preprint arXiv:1806.03713(2018).
- [4] Chen, Yongheng, Chunyan Yin, and Wanli Zuo., "Multi-task learning for stance and early rumor detection", Optical Memory and Neural Networks 30.2: 131-139(2021).