ユーザ信頼性を考慮したマルチタスク学習によるデマ検知

Lijing Qu† 成 凱†‡

†九州産業大学大学院情報科学研究科 データサイエンス・人工知能領域

〒813-8503 福岡県福岡市東区松香台 2-3-1

E-mail: †‡chengk@is.kyusan-u.ac.jp

あらまし SNS 上で発信された情報の中に，正しい情報だけでなく，デマを含む偽情報も拡散され，社会問題となっている．デマ検知のためにはユーザ属性，投稿コンテンツと拡散ネットワーク等複数の側面からデマの特徴を抽出し，検知を行うことが一般的である．また，単一側面ではデマ検知の精度が低く，複数の側面を用いたマルチタスク学習が必要とされている．しかし，先行研究では一部の側面しか用いられなかった．本研究ではユーザ信頼性を考慮したスタンス分類と極性分類を補助タスクとする分類モデル提案し，SemEval-2019 Task7 で公開される Twitter データセットを利用し，デマ検知の精度を評価した．

キーワード 情報信憑性，デマ検知，感情分析，深層学習，マルチタスク学習

1. 初めに

インターネットの普及に伴いソーシャルメディアをはじめとする便利なプラットフォームが急増し，誰でも自由に発信できるようになっている．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，政治的・金銭的利益のもとで発信された偽情報や真偽不明の情報も拡散されており，健康に深刻な被害をもたらしたり，社会的混乱が生じたりする等，社会問題となっている[1][16][17]. 偽情報の中に，特にデマと呼ばれるものが存在し，社会情勢が不安な時などに発生して人心を惑わすような憶測や事実誤認による情報や単なる悪口や根拠のないウワサ話，流言飛語もデマの範囲になっている[13][14][15]．不確かな情報や根拠のない推測，個人の思い込み，悪意のある虚偽情報などが簡単にネット上へと発信され，これらの偽情報の一部は，ツイッターで他の人の投稿を再投稿する「リツイート」などの SNS の機能を使って，短時間で次から次へと拡散される．内容が衝撃的であればあるほど，義務感や正義感などからその情報をさらに世の中に広めようとする傾向があるので，迅速に対処することが大事である[5][8][18].

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．デマ検知のためにはユーザ属性，投稿コンテンツと拡散ネットワーク等複数の側面からデマの特徴を抽出し，検知を行うことが一般的である．近年，深層学習と感情分析がデマ検知に使われている[5][10]．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とされている．

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．感情分析を用いたデマ検知では，投稿コンテンツの感情極性または感情辞書，ルールによる算出した感情スコアを特徴量の一つとして扱うことが多い[8][9]. また，スタンス分類を補助タスクとして，メインタスクであるデマ検知と共同学習する方法もある[7]．しかし，投稿コンテンツのテキスト解析による感情極性の分類精度が低く，それに基づくデマ検知の誤検出の割合が高いという問題が指摘されている．また，ユーザのスタンス分析と極性分析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．

そこで本研究では深層学習に基づく感情分析（深層感情分析）を用いてソース投稿の極性分析と返信投稿のスタンス分析を行うことでデマ検知の精度を高める手法を提案する．マルチタスク学習を用いてデマ検知をメインタスクとし，ユーザ信頼性を考慮したスタンス分類と極性分類を補助タスクとする深層分類モデルを構築し評価実験による検証を行う．

2. 関連研究

機械学習に基づくデマ検知は現在主な手法となっている．Qazvinian ら[1]はツイッターのデマ情報に対し，テキストコンテンツ，ユーザ行為とマイクロブログにおける特徴量を使用し，デマ検知実験を行なった．Kwon ら[2]はデマ情報拡散の周期性，構造性とテキストコンテンツの言語性から分析し，決定木，ランダムフォレストと SVM 分類機でデマを検知する実験を行った．Vosoughi ら[4]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．

また，感情分析を用いたデマ検知に関する研究も多

数存在する．須田ら[16]は感情極性辞書を用いて収集したツイートの感情極性値を算出し，特徴量の一つとしてデマ識別実験を行なった．Bhutani ら[8]はデマ情報とニュース情報の感情極性特徴を学習し，作成したモデルはテストデータの分類を予測した．Azri らは画像コンテンツと感情表現を注目し，deepMONITOR を開発した[10].

マルチタスクに基づくデマ検知に関する研究も行っている．Kochkina ら[4]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．また，Chen ら[11]はスタンス分類タスクと早期デマ検知タスクの共同学習を行った．

3. 偽情報の一つとしてのデマの対策

デマは偽情報の一種類として本来，政治的な目的で相手を誹謗し，相手に不利な世論を作り出すように流す虚偽の情報である．現在では，社会情勢が不安な時などに発生して，人心を惑わすような憶測や事実誤認による情報や単なる悪口や根拠のないうわさ話，流言飛語もデマの範囲になっている．デマの真実性が未検証，または偽であるステートメントと定義されている．さらに，偽である場合，一部の研究ではそれを「フェイクニュース」と「虚伝」と細分化されていた．

3.1. デマの特徴

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．デマの発信者は人々の気持ちを煽って，行動を起こさせるという悪意に近い意図があるため，事実と反する情報を，いち早く否定するにはわかりやすさが重要である．特に，SNS 上で拡散されたデマは特定の状況を狙って発信し，インサイダー形式で情報正当化を強調する特徴がある．例えば，「テレビ局のプロデューサーからの情報:政府が来月 1 日に緊急事態宣言を出し，２日にロックダウン＝都市封鎖を行う」という，「緊急事態宣言」という特定の状況を狙って，「テレビ局のプロデューサーからの情報」で情報正当化を強調している．また，デマは特定の雰囲気を作成し，急速に強い感情を誘発する場合が多い．例えば，「ヤバイ東京封鎖宣言で，パニック状態に突入！スーパーがとんでもない混雑になり水・食料品が買い占めで消える.」というデマには，「ヤバイ」，「とんでもない」のような言葉がよく書かれて背後に民衆を恐怖させ，焦慮させ行動を起こさせる悪意に近い意図が伺わせる．

3.2. デマ対策

デマ対策はデマ追跡，デマ検知，スタンス分類，真実性検証という複数のタスクから構成される．**デマ追跡**はデマが発信された時点から，訂正されるまでの情報と意見を収集することである．**デマ検知**はある投稿がデマかデマでないかの検出と指す．主に通常の投稿，意見，冗談ではないかの判断である．**スタンス分類**はソース投稿（最初のメッセージ）に対するユーザの反応の分類と指す．主に支持，否定，疑いとコメントという四つのカテゴリで使われている．**真実性検証**はデマ検知で抽出したデマは本当にデマかどうかの判断と指す．

デマ検知の基本手法としては主に三つに分けられる．一つ目はユーザ情報を利用して発信者の信頼性を分析し，発生源からデマを特定する．二つ目は拡散のメガニズムとネットワークを解析し，デマの特有なパターンを利用し，デマを特定する．三つ目はコンテンツ情報やそれに対する反応を対象とし，テキスト解析や画像解析などを行い，デマを特定する．

A. ユーザ情報に基づく手法

主に SNS 上で登録された統計的な情報と指す．プラットフォームにより状況が多少異なる．共有的な特徴量はユーザ名、登録時間、投稿数、友達数、プロフィル、端末種類、地域などが挙げられる。

B. 拡散ネットワークに基づく手法

ネットワークのタイプに関しては Kai Shu ら[5]が同種型と異種型に分類される．同種型のネットワークは(a)ユーザの関係ネットワーク(Friendship Network)，(b)拡散ネットワーク(Diffusion Network)と(c)信頼性ネットワーク(Credibility Network)が存在する[5]．異種型ネットワークは(d)知識ネットワーク(knowledge Network)、(e)スタンスネットワーク(Stance Network)と(f)相互作用ネットワーク(Interaction Network)が存在する。

C. コンテンツ情報に基づく手法

欺瞞スタイル理論によれば，読者を欺くことを目的とした欺瞞情報のコンテンツスタイルは，たとえば誇張された表現や強い感情を使用するなど，真実のコンテンツスタイルとは多少異なる必要がある．デマに関するコンテンツ情報はテキストコンテンツと視覚的コンテンツが分けられる．

テキストコンテンツはソース投稿とソース投稿に対する反応を含まれている．これまでのデマ検知に関する研究はほとんどテキストコンテンツを利用されている．テキストコンテンツのソース投稿以外に，一般ユーザの返信またはコメント情報から得られた嘘に対する意見や態度もデマ検知に重要である．

視覚的コンテンツは画像，ビデオなどが含まれている．悪意で作成された画像・ビデオは人々を誤解させる可能性があるので，画像・ビデオの鮮明さ，一貫性，多様性などを考慮した研究が行われている．

3.3. 深層感情分析に基づくデマ検知

デマに限らず,偽情報を拡散させるのは人間であり,人間の主観的判断に左右されることが多い．従って,発信者の意図や信頼性を分析しデマ検知に適用することが重要である．

3.3.1. デマ発信者の意図とユーザの信頼性

デマの発信者の意図を汲み取るためには社会背景,政治面，金銭面，利益面，心理面など様々な方面から分析する必要がある．近年，ネット有名人になりたい人が増えてきて，フォロワーを増やすため，注目度を高めるために真偽を判断せずに目を引くものであれば,SNS にデマを発信してしまう．

それはただ自身の心理満足であるが，悪い意図ではない．悪意を持ってある人間，組織または国を中傷するために拡散されるデマは大きな影響を及ぼしている．例えば,デマによって企業のイメージ低下だけでなく,売り上げや価格の低下，株価の暴落を引き起こす事例は数多いものであり，採用活動や実務に影響が出ることがある．

デマ発信者以外のユーザも全て信用できるわけではない．特に，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．信頼性の高いユーザとは政府機関,専門家,社会地位の高い方であり,信頼性の低いユーザは実名検証されていいないアカウント，以前デマを発信・拡散したことのあるユーザである．

「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．

3.3.2. 意図を汲み取るための感情分析

感情分析(Sentiment Analysis)は意見マイニング(Opinion Mining)とも呼ばれ,テキストで表現されたエンティティとそれらの属性に対する人間の意見(opinion)，感情(sentiment)，感覚，評価と態度を分析する研究分野と指す[9].

自然言語処理(Natural Language Processing)のミニタスクとして，初期の研究ではテキストの評価極性（ポジティブ，ネガティブ，ニュートラル）を判定する研究を始めた．その後，詳細なカテゴリに分類する研究が行われている．現在，感情分析という技術は意見抽出，意見マイニング，極性判定，立場分析の総称となっている．

意見分析はますます重要になってきた．なぜなら,意見は人間の行動に影響を与える重要な要素であり,決断する際に他人の意見を求める場合もよくあり，人間の信念や認識は他人の意見に左右される可能性がある．例えば,EC サイトで商品を購入する際に，レビュー評価を参考する．

意見(Opinion)はある対象に対する個人的判断と評価と定義する．一般的に，エンティティ(entity)，アスペクト(aspect)，感情(sentiment)，意見保有者(holder),時間(time)という五つの要素から構成される．エンティティは評価の目標対象，アスペクトはエンティティの属性，感情は具体的な評価と感覚，意見保有者は意見の発信者，時間は意見の発信時点とのことである．

感情(Sentiment)は意見に関連する基本的な感情，態度，評価，感覚と定義する．一般的に，タイプ(type),極性/傾向(orientation)，強度(intensity)という三つの要素から構成される．タイプは理性的・客観的また非理性・主観的に分類され,例えば,「この車は値段の価値がある」という文は理性的な発言とし,「これは最高の車」は非理性・主観的な発言と分類される．極性/傾向はポジティブ，ネガティブまたはニュートラルとのこととさす．強度はポジティブまたはネガティブだと感じる強さであり,「非常に」,「あまり」などの形容詞が挙げられる．

3.4. マルチタスク学習

デマ情報を分類・推定するには，最も用いられている手法は機械学習である．しかし，前に述べた通り,デマの特徴はユーザ属性，投稿コンテンツと拡散ネットワーク等複数の側面から理解する必要がある．単一側面ではデマ検知の精度が低く，複数の側面を用いたマルチタスク学習が必要であるとされている[4][7].

あるモデルによって得られた出力が次のモデルの入力となる時，これらのモデルが独立であれば，このようなシステムはパイプライン(pipeline)システムと呼ばれる．パイプラインと対比される別のアプローチとして,モデルのカスケード接続(model cascading)がある．

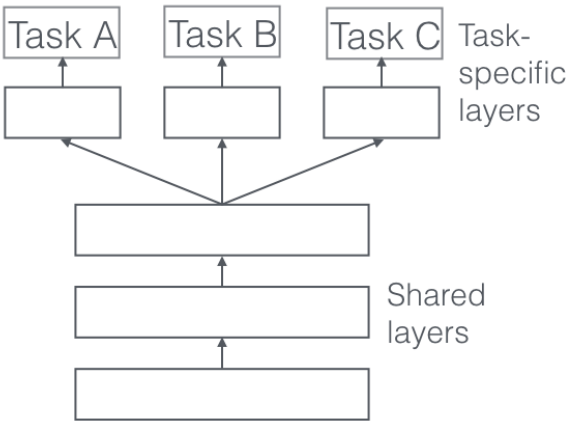


図 1 マルチタスク学習

モデルのカスケード接続に関連する技術としては図 1 に示すようなマルチタスク学習(Multi-Task

Learning：MTL)があり，深層学習システムにおいて容易に実装可能である．マルチタスク学習はメインであるタスクに関係がある複数のタスクを一つのモデルに同時に学習を行わせることで精度を向上させる手法である．自然言語処理の例を挙げると，文から誰が，誰に対して，何をしたかといった情報を抽出する問題を解くために，各単語の品詞を当てるタスクを同時に解く[7]．マルチタスク学習においては複数の関連する予測タスクを取り扱うが，これらの予測結果はタスク間で入力関係にある場合もない場合もありうる．

異なるタスク学習に対してはそれぞれのネットワーク割り当てるのが通常であるが，マルチタスク学習ではネットワーク間で構造やパラメータの一部を共有させる．このようにすると，予測において共通する部分（共有構造）は全てのタスクから影響を受けることになるので，あるタスクに対する訓練データが他のタスクにおける予測の改善を促進する可能性がある．

4. 評価実験

SemEval-2019 Task 7：RumorEval 2019:Determining Rumor Veracity and Support for Rumors で提供されたデータセットを使用する．SemEval のタスクは，多くの種類の意味注釈(Semantic Annotation)を設けており，それぞれ様々なスキーマを持っている．データセットに 7 種類異なる話題に関連する Twitter の会話スレッド(Conversation threads)を採用した．

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．データセットのラベルに関してはデマの真実性を解決する公式声明または他の信頼できる証拠源を特定するチームの報道記事の記者および専門家メンバーによって手動で付けたものである．

4.1. データセットの統計分析

デマの特徴等を理解するために，データセットに対してユーザ関係，ユーザスタンス，感情極性に対して統計分析・可視化を行った．データ可視化は数値だけで確認しにくい現象や関係性，変化性などを一目見ればわかる形（可視化グラフ，チャート，表，画像等）に変換し，データに隠した情報を表示して，数字から分かる情報の理解を助けることである．今回は投稿種別（真実，デマ，未検証）によって，ユーザ関係ネットワーク，ユーザスタンス割合，感情極性割合を統計的に分析し，可視化を行った．

4.1.1. ユーザ関係ネットワーク

発信者のユーザ関係を理解するために，Gephi というライブラリを使って，ユーザ関係ネットワークを作成した．Gephi は高彩度の美しいグラフを直感的な操作で作成することができ，特に様々なコミュニティ組織やスモールワールドネットワークをマップ化できるソーシャルデータコネクタを簡単に作成することができるため，ソーシャルネットワーク分析によく使われている．ネットワークは点（Node）とそれらを繋ぐ線（Edge）からなる．Node はユーザ ID，Edge はフロー，転送，返信等の関係と示す．

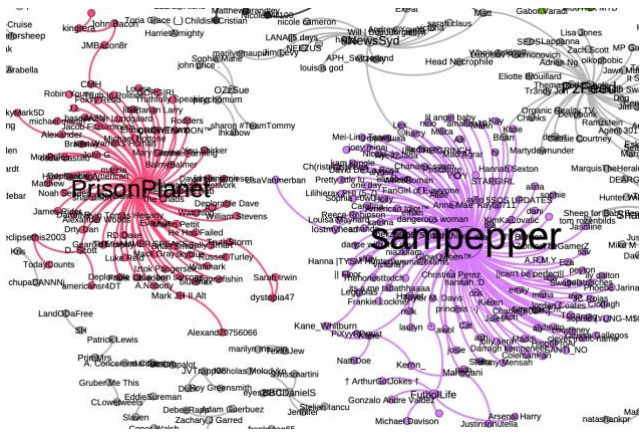


図 2 デマデータのユーザ関係ネットワーク

図 2 はデマデータのユーザ関係ネットワークを示しており，左部分は「PrisonPlanet」に関するユーザ関係ネットワークであり，右部分は「sampepper」に関するユーザ関係ネットワークである．PrisonPlanet はデマデータの発信者のユーザ名であり，同時に PrisonPlanet ネットワークのオピニオンリーダーである．同様に，sampepper も sampepper ネットワークのオピニオンリーダーである．特に，他人を誹謗中傷するデマデータに対して，多くのユーザが議論に参加し，批判または反論することを招く様子がうかがわせる．また，二次拡散の様子も確認できる．

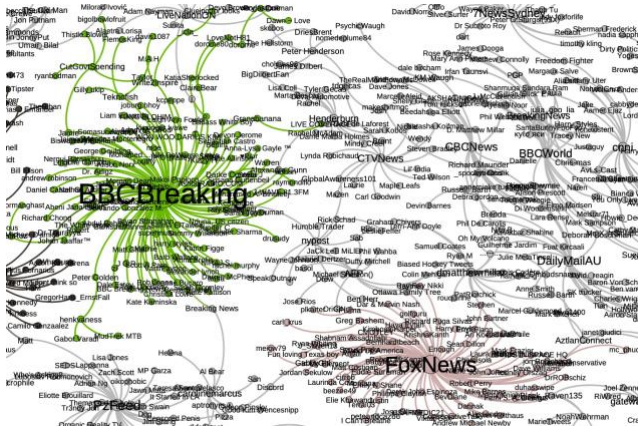


図 3 非デマデータのユーザ関係ネットワーク

図 3 は非デマデータのユーザ関係ネットワークを示しており，左部分は「BBCBreaking」に関するネットワークであり，右部分は「FoxNews」に関するネットワ

ークである．BBCBreaking はニュース情報の発信者のユーザ名であり，同時に BBcBreaking ネットワークのオピニオンリーダーである．同様に，FoxNews も FoxNews ネットワークのオピニオンリーダーである．ニュース情報のユーザ関係ネットワークはデマデータのユーザ関係ネットワークの形はほぼ一緒である．二次，三次拡散する様子を見える．

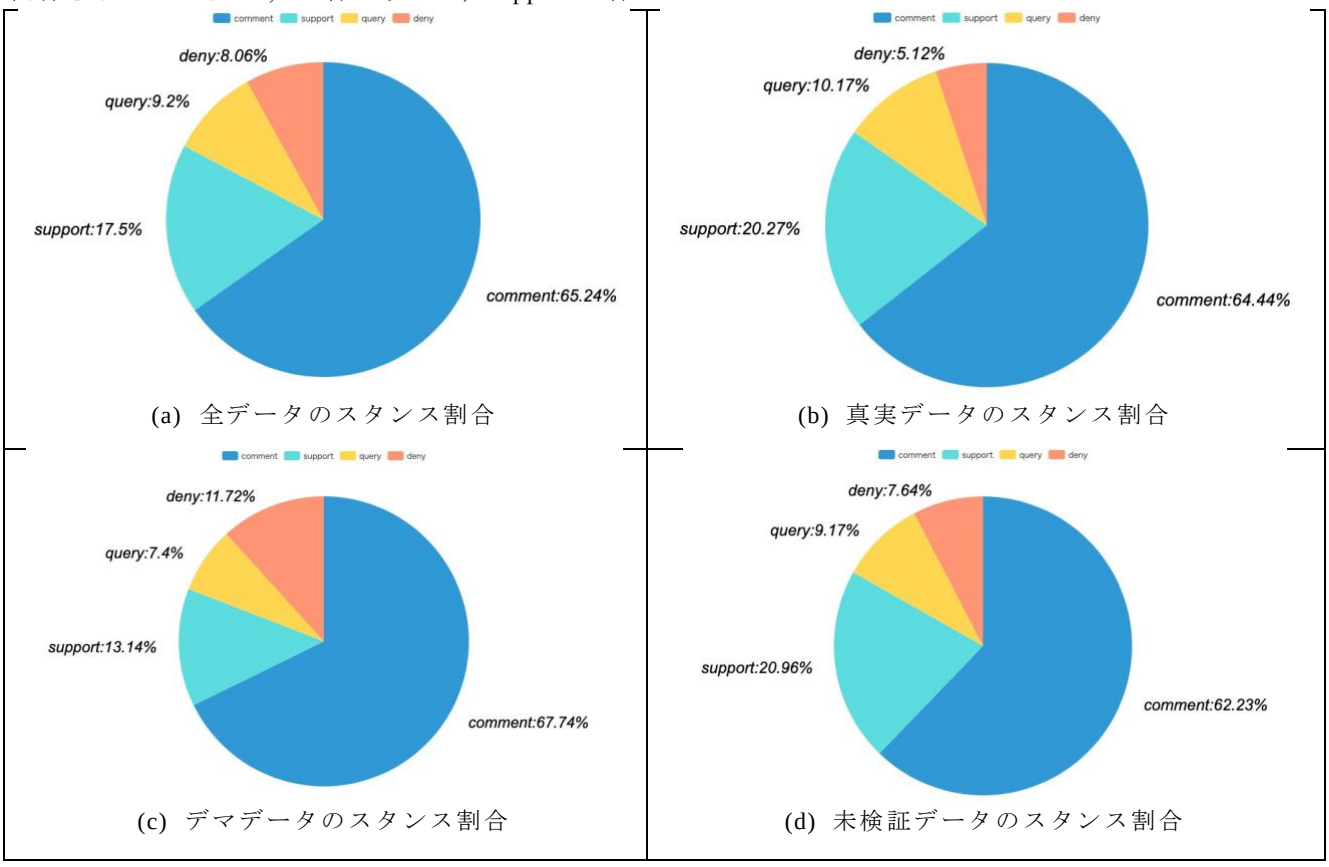
4.1.2. スタンスの割合

ユーザスタンス（support:支持,deny: 反対, query: 質疑, comment:コメント）の割合を解析した．結果は図 4 にまとめている．

図 4 (a)によると，全てのデータ 3,228 件のうちに，support の件数は 565 件，全体の 17.5%を占める．query の件数は 297 件，全体の 9.2%を占める．deny の件数は 260 件，全体の 8.06%を占める．comment の件数は 2,106 件，全体の 65.24%を占める．つまり，ある情報に対して，6 割以上のユーザは単純にコメントを示している．また，query と deny の割合の合計は 17.26%であり，support の割合は 17.5%である．疑う，否定する態度を持つ人と支持する態度を持つ人の割合はほぼ同じである．

図 4 (b)は真実性が真である「真実データ」のスタンス割合を示している．1,485 件のうちに，support の件数は 301 件，全体の 20.27%を占める．query の件数は 151 件, 全体の 10.17%を占める. deny の件数は 76 件，全体の 5.12%を占める．comment の件数は 957 件，全体の 64.44%を占める．query と deny の割合の合計は 15.29%であり，support の割合は 20.27%である．つまり，疑う，否定する態度を持つ人の割合と比べると，支持する態度を持つ人の割合は高いことがわかる．また，真実性が真のデータに対して，否定的な態度を持つ人の割合は 5.12%しかないため，ほとんどのユーザは真実に否定しない．

図 4 (c)は真実性が偽である「デマデータ」のスタンス割合を示している．デマデータ 1,271 件のうちに，support の件数は 167 件, 全体の 13.14%を占める. query の件数は 94 件，全体の 7.4%を占める．deny の件数は 149 件, 全体の 11.72%を占める. comment の件数は 861 件，全体の 67.74%を占める．support の割合の合計は 13.14%であり，query と deny の割合は 19.22%である．図 4 (b)で示した真実データの状況と違い, 支持する態度を持つ人の割合と比べると，疑い，否定する態度を持つ人の割合は高かった．特に，デマに支持する人の割合 13.14%とデマに否定する人の割合 11.72%は数字的に近い．それは偽情報に対して，ユーザ間の激しい議論を引き起こし，反論する人が多い．



(a) 全データのスタンス割合　(b) 真実データのスタンス割合

(c) デマデータのスタンス割合　(d) 未検証データのスタンス割合

図 4 ユーザスタンスの解析結果

図 4 (d)は真実性が未検証である「未検証データ」のスタンス割合を示している．未検証データ 3,228 件のうちに，support の件数は 301 件，全体の 20.96%を占める．query の件数は 151 件，全体の 9.17%を占める．deny の件数は 76 件，全体の 7.64%を占める．comment の件数は 957 件, 全体の 62.23%を占める. query と deny の割合の合計は 16.81%であり，support の割合は 20.97%である．未検証データに対して，疑う，否定する態度を持つことより，ユーザは信じることを選択する傾向がある．

4.1.3. 感情極性の割合

図 5 はソース投稿の感情極性に関する統計結果を示している．

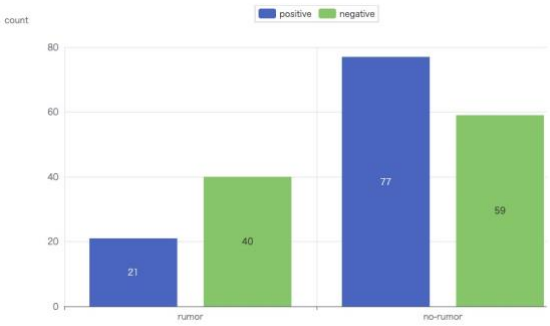


図 5 ソース投稿の感情極性

これによると，デマのソース投稿 61 件のうち，テキスト解析によるポジティブな投稿は 21 件，ネガティブな投稿は 40 件である．ポジティブな投稿 : ネガティブな投稿は約 1:2 になり，デマのソース投稿の中でネガティブな表現はポジティブな表現より多いことが分かった．

デマでないデータ（真実データと未検証データ）136 件のうち, テキスト解析によるポジティブな投稿 77 件, ネガティブな投稿 59 件である．デマでないデータの中でポジティブな表現はネガティブ無表現より多いことがわかった．

4.2. マルチタスク学習によるデマ検出

4.2.1. データの前処理

データの前処理は形態素解析，不用語削除，ラベル変換，データ分割という一連の操作から構成された．形態素解析では，Python の自然言語処理用ライブラリ NLTK を利用した．具体的に，NLTK の nltk.word_tokenize を使って英語の単語を分割した．不用語削除では，「^¥%*$#~, . /|-_=+→←⇨⇦」などの符号を削除し，「!」や「?」など感情を判断するために使える符号は保留されていた．

ラベル変換では, スタンス分類の支持, 否定, 疑問, コメントを「0」,「1」,「2」,「3」に変換し，感情分析のポジティブ，ネガティブを「1」,「2」に変換し，デマ検知の真実性の真，偽，未検証を「0」,「1」,「2」に変換した．

データ分割では合計 2,339 件のデータをトレニングデータ 1754 件とテストデータ 585 件に分割した．分割したデータ数は表 1 のように示す．

表 1 訓練とテストデータの分割

	Train	Test	Data
0（真）	836	260	1096
1（偽）	701	237	938
2（未検証）	217	88	305
Total	1754	585	2339

4.2.2. パラメータ最適化

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)手法[7]を使用し，hyperopt で実装した．パラメータの設定と最適化した結果は表 2 に示す．

表 2 パラメータ最適化

パラメータ	オプション	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	100	100
LSTM 層の数	{1,2}	**1**
LSTM 層のユニットの数	{100, 200, 300}	**300**

4.2.3. モデル構築

シーケンスデータを処理する際に，異なる長さを持つことは一般的である．各会話に含むツイッター数も異なるし，各投稿の返信数も異なる．前処理でベクトル化したデータの長さを一致するため，Masking 層を追加した．

Shared LSTM 層はハードパラメーター共通層として, その後に複数のタスク固有の層が続いた．各タスクでは, 最適化したパラメータを採用し, 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 keras で構築した．また，スタンス分類タスクと感情分析タスクは Twitter レベルなので, 共有層 LSTM 以外に LSTM 層を追加しないことにした. マルチタスクのモデルは図 6 のように示す．

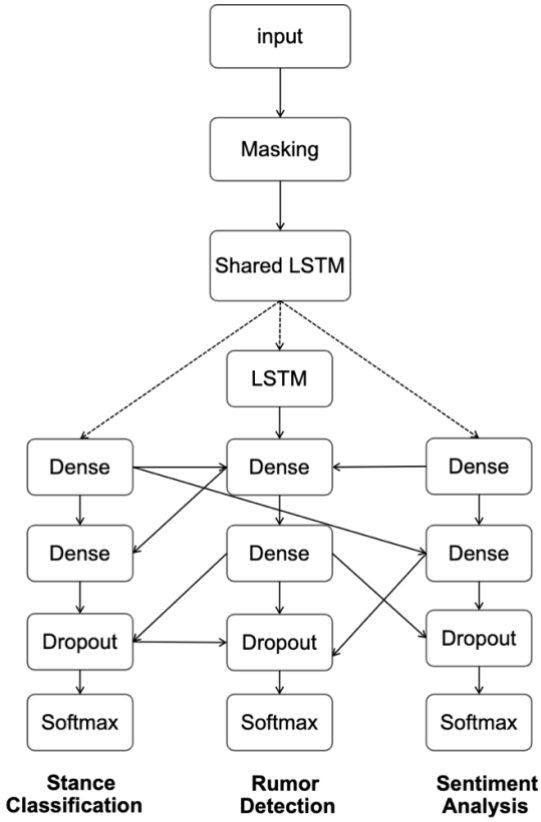


図 6 マルチタスクのモデル

各 timestep で同じ処理を行うため，TimeDistributed を使った．ニューラルネットワークの過学習を防ぐために，Dropout 層を追加した．Dropout は一定の確率でランダムにニューロンを無視して学習を進める正則化の一種類として，複数の構造を持つニューラルネットワークを個別に学習させ，各分類器の出力の平均値を認識結果とする．最後に，活性化関数 Softmax を出力層として構築した．

4.2.4. 評価指標

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．

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

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．

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

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

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

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．

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

なお，nはクラスの数，$F1_i$は各クラスの$F1$.

4.2.5. 結果と考察

表 3 と表 4 は深層感情分析を使わない分類結果と深層感情分析を使用したマルチタスク学習による分類結果をそれぞれ示している．

表 3 深層感情分析を使わない単一タスク学習結果

	Precision	Recall	F1-score
0 (真)	0.45	0.97	0.62
1(偽)	0.31	0.02	0.03
2(未検証)	0.33	0.02	0.04
Accuracy			0.45
Macro avg	0.37	0.34	0.23
Weighted avg	0.38	0.45	0.30

表 4 深層感情分析を使うマルチタスク学習結果

	Precision	Recall	F1-score
0 (真)	0.72	0.71	0.72
1(偽)	0.69	0.75	0.72
2(未検証)	0.51	0.42	0.46
Accuracy			0.69
Macro avg	0.64	0.63	0.63
Weighted avg	0.68	0.69	0.68

深層感情分析を使わない単一タスク学習では分類正解率は 45%であり，真と偽の F1 は 0.62 と 0.03 であった．それに対して，深層感情分析を使ったマルチタスク学習の分類正解率は分類の正解率は 69%であり，True と False 情報の F1-score は 0.72 となる．7 割のデータが正しく分類され，ある程度でデマを識別できたと考えられる．

5. 終わりに

本研究では，深層感情分析を用いたマルチタスク学習によるデマ検知を提案し,実験による評価を行った．テキスト解析による感情極性の分類精度が低く，それに基づくデマ検知の誤検出の割合が高い点とスタンス分類にユーザ信頼性を考慮していないため，本研究では深層感情分析を用いて，ユーザ信頼性を考慮した感情極性分類とスタンス分類がデマ検知と共同学習する手法を提案した．スタンスと感情極性がデマ検知に有効ではあるが，ユーザ関係ネットワークはデマと非デマははっきり区別できないことを判明した．

データ解析の実験では，スタンスと感情極性がデマ検知に有効ではあるが，ユーザ関係ネットワークはデ

マと非デマははっきり区別できないことを判明した．分類実験では，7 割のデマデータが正しく分類され，ある程度はデマを識別できたと検証した．それによって，深層感情分析がデマ検知に有効であることがわかった．

今後の課題としては分類精度をさらに高めることが挙げられる．デマは歴史，文化，集団心理等の要素と深く関わるため，そういった要素をいかに学習モデルに反映させるか，今後の課題となる．

謝辞

本研究の一部は，KSU 基盤研究費 K060069 の助成を受けたものである．

参 考 文 献

[1] Vahed Qazvinian, et al, (2011). “Rumor has it: Identifying misinformation in microblogs”, In Proceedings of the 2011 Conference on Empirical Methods in Natural Language Processing, pp. 1589-1599.

[2] Kwon, Sejeong, et al, (2013). “Prominent features of rumor propagation in online social media.” ,2013 IEEE 13th international conference on data mining, pp. 1103-1108.

[3] Soroush Vosoughi. (2015). "Automatic Detection and Verification of Rumors on Twitter ". Ph.D. Dissertation, Massachusetts Institute of Technology 2015.

[4] Kochkina, E., Liakata, M., and Zubiaga, A. (2018). "All-in-one: Multi-task learning for rumour verification". In Proceedings of the 27th International Conference on Computational Linguistics. 2018, pp. 3402–3413

[5] Kai Shu, H Russell Bernard, and Huan Liu (2018). “Studying Fake News via Network Analysis: Detection and Mitigation”. Emerging Research Challenges and Opportunities in Computational Social Network Analysis and Mining. Lecture Notes in Social Networks. Springer, Cham, pp. 43–65. https://doi.org/10.1007/978-3-319-94105-9_3

[6] Arkaitz Zubiaga, Elena Kochkina, Maria Liakata, Rob Procter, Michal Lukasik, et al. (2018b). “Discourse-aware rumour stance classification in social media using sequential classifiers”. Information Processing & Management, Volume 54, Issue 2,2018, pp. 273-290.

[7] Sener and Koltun. (2018). “Multi-Task Learning as Multi-Objective Optimization”. In Proceedings of the 32nd International Conference on Neural Information Processing Systems (NIPS'18). December 2018, pp. 525–536.

[8] Bhutani, Bhavika, et al, (2019). "Fake News Detection Using Sentiment Analysis," 2019 Twelfth International Conference on Contemporary Computing (IC3), Noida, India, 2019. pp. 1-5, doi: 10.1109/IC3.2019.8844880.

[9] Bing Liu (2020). Sentiment Analysis: Mining Opinions, Sentiments, and Emotions, Cambridge University Press; Second Edition (2020/10/15).

[10] 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). pp. 497–513

[11] Chen, Yongheng, Chunyan Yin, and Wanli Zuo. (2021). “Multi-task learning for stance and early rumor detection”, Optical Memory and Neural Networks 30.2: 131-139.

[12] Yoav Goldberg(2019) ,「自然言語処理のための深層学習」，共立出版株式会社 2019 年 1 月.

[13] 齋藤勇,「イラスト図解 デマの心理学 怖い群集心理のメカニズム」，宝島社 (2020/7/10)

[14] シナン・アラル(著), 夏目大 (翻訳),「デマの影響力 なぜデマは真実よりも速く，広く，力強く伝わるのか?」，ダイヤモンド社 (2022/6/8)

[15] 物江潤,「デマ・陰謀論・カルト」，新潮社 (2022/11/17)

[16] 須田剛裕等 (2013).「震災時におけるツイッターのトレンドワードと拡散情報を利用したデマ推定の一考察」，情報処理学会第 75 回全国大会講演論文集 2013.1: 99-100.

[17] 総務省(2020).「プラットフォームサービスに関する研究会 最終報告書」，https://www.soumu.go.jp/main_content/000668595.pdf (2022 年 11 月 19 日アクセス)

[18] 日経新聞社,「フェイクニュース，ＳＮＳで拡散し社会に混乱」，経済ナレッジバンク・ビジュアル・ニュース解説 2020.7.20