ニュースコンテンツとソーシャルコンテクストを用いたフェイクニュースの早期自動検出

谷　聡馬† 佐々木裕多† 張　建偉†

† 岩手大学理工学部　〒 020–8551 岩手県盛岡市上田 4-3-5
E-mail: †{s0619039,s0619027,zhang}@iwate-u.ac.jp

あらまし　近年の SNS の発達に伴い，フェイクニュースの拡散が問題となっている．フェイクニュースは，政治・経済をはじめとした様々な側面で社会に悪影響を及ぼす危険性があり，拡散初期段階での早期自動検出技術が求められている．本研究では，ニュースコンテンツとそれに関連するツイートデータを用いた，機械学習による 2STEP のフェイクニュースの早期検出手法を提案する．STEP1 では，ニュース記事の発行直後を想定し，ニュースコンテンツのみを用いてフェイクニュースの早期検出を図る．STEP2 では，ニュース記事の拡散段階を想定し，STEP1 で真偽を判別できなかったニュースコンテンツに対し，ソーシャルコンテクストを時系列に沿って付与することにより真偽を判別していく．両方の STEP において，真偽の予測確率が閾値以上のニュースを追跡対象から取り除くことで，検出の精度と早期性の両立を図る．SNS ユーザーのデータをソーシャルコンテクストとして用いた実験では，提案手法を用いることにより，ニュースの拡散から 5 分後と 10 分後のタイミングにおいて，ベースラインよりも精度と F1 値が向上することが確認された．リプライツイートを用いた実験では，提案手法による精度と F1 値の向上は見られなかったものの，リプライツイートを用いることがフェイクニュースの検出率の向上に役立つことが確認された．

キーワード　テキストマイニング，言語モデル，フェイクニュース検出，ソーシャルコンテクスト

1 はじめに

近年，インターネットやスマートフォンの普及に伴い，ソーシャルメディアやソーシャルネットワークサービス（以下 SNS）の利用者が増加している．これらのプラットフォームは有益な情報源として利用可能な反面，悪意をもって作成された事実と異なる内容のニュース（以下フェイクニュース）の拡散によって，社会に悪影響を及ぼす原因にもなりうる [1] [2] [3]．ソーシャルメディアの利用者は，SNS 上にニュース記事の共有や投稿を行うことが可能である．これによりユーザー同士の議論が活発になることが，フェイクニュースの拡散を加速させる一因となっている [4]．2019 年 12 月以降の新型コロナウイルスの世界的な流行においては，新型コロナウイルスに関するフェイクニュースがソーシャルメディアや SNS を通して拡散された [5]．

フェイクニュースの判別には専門家によるファクトチェックが有効であるが [6]，ソーシャルメディアや SNS 上の無数のニュースのファクトチェックを全て人手で行うことは非常に困難である．また，フェイクニュースは真実のニュースよりも拡散の速度と規模が大きいという特徴を持つ [4]．そのため，フェイクニュースを自動的かつ早期に検出するための技術が求められ，様々な研究が行われている [7] [8] [9]．

フェイクニュース検出に用いられる特徴量は，ニュースコンテンツとソーシャルコンテクストに大別される [10] [11]．ニュースコンテンツとは，ニュースのタイトルや本文など，ニュースそのものから得られる情報のことである．一方，ソーシャルコンテクストとは，ニュースの拡散経路や拡散に関わったユーザーの情報など，ニュースの拡散に関わる情報を指す．最新の研究では，両方の特徴量を用いることによる精度の向上が報告されている [12]．しかし，十分なソーシャルコンテクスト特徴量を得るためには，ニュースの拡散を待つ必要があるため，検出の精度と早期性のトレードオフが問題となる．

我々は，ニュースコンテンツとソーシャルコンテクストを特徴量として用いた，機械学習による 2 段階（以下 STEP1, STEP2）のフェイクニュースの早期検出手法を提案する．STEP1 では，ニュースの発行直後を想定し，ニュースコンテンツのみを用いてフェイクニュースの早期検出を図る．STEP2 では，ニュースの拡散段階を想定し，STEP1 で真偽を判別できなかったニュースコンテンツに対し，ソーシャルコンテクストを時系列に沿って付与することにより真偽を判別していく．両方の STEP において，真偽の予測確率が閾値以上のニュースを追跡対象から取り除くことで，検出の精度と早期性の両立を図る．提案手法を用いることにより，ニュースの拡散初期において，ベースラインと比較して精度と f 値の向上が確認された．

2 関連研究

2.1 フェイクニュース検出に関する関連研究

Zhou ら [4] は，フェイクニュースに関連する基礎理論について分野を跨いだ調査を行った．また，複数の異なる観点によるフェイクニュース検出手法を評価し，今後の研究課題について示した．

山中ら [13] は，Twitter 上のイベントに対して，イベントを構成するツイートの特徴量を作成し，古典的な機械学習モデル

を用いてイベントの真偽を時系列に沿って判別した．この研究では，本研究と同様にモデルを 2 段階に分け，一定の条件を満たしたイベントの予測ラベルを断定することで，イベントの発生規模を考慮した偽情報検出の手法を提案した．本研究では，フェイクニュースを検出対象とし，テキスト特徴量のみを使用している．また，古典的な機械学習モデルに加えて深層学習モデルも比較対象として用いている．

Raza ら [12] は，事前学習済み言語モデルである BART [14] を用いた，ニュースコンテンツとソーシャルコンテクストを利用したフェイクニュース検出手法を提案した．ニュースコンテンツとして NELA-GT-2019 [19] データセットを，ソーシャルコンテクストとして Fakeddit [20] データセットを利用した．また，転移学習を用いてユーザーの信頼性を推定する手法や，複数の特徴量を組み合わせることで弱ラベルからニュース記事のラベルを推定する手法について提案した．比較実験により，BART のエンコーダ部分とデコーダ部分，およびニュースコンテンツ特徴量とソーシャルコンテクスト特徴量がそれぞれ検出精度の向上に貢献していることを示した．また，提案手法により拡散初期段階から高い精度を実現できることを示した．本研究では，2STEP の検出手法により，フェイクニュースの早期検出を図る．

2.2 BART

BART（Bidirectional Auto-Regressive Transformer）[14] は, 事前学習済み Transformer ベースモデル [15] であり, BERT（Bidirectional Encoder Representations from Transformers）[16] を GPT（Generative Pre-trained Transformer）[17] と組み合わせることで，seq2seq タスク（入出力ともにシーケンスのタスク）を可能にした．BERT は双方向エンコーダとして用いられ，ランダムにマスクされた単語を予測することで学習を行う．GPT は自己回帰性デコーダとして用いられ，入力された単語から次の単語を予測することで学習を行う．BART における文章分類タスクでは，エンコーダとデコーダ双方に同じ文章を入力し，末尾のトークンの最終層の用いて分類を行う．

3 提案手法

本研究では，特定のニュース記事がフェイクニュースであるか，真実のニュースであるかを判別する．提案手法の全体像を図 1 に示す．本手法において，i 番目のニュース記事を N_i，陽性の予測確率を $P(pos)$，陰性の予測確率を $P(neg)$，i 番目のニュースが陽性または陰性と予測された確率をそれぞれ $P(pos|N_i)$，$P(neg|N_i)$ とする．$P(pos|N_i)$，$P(neg|N_i)$ は $P(pos|N_i)+P(neg|N_i)=1$ を常に満たす．また，ニュースの拡散開始からの経過時間を $t_n(n=0,1,2,\ldots)$，陽性または陰性のラベルを断定するための予測確率の閾値を，それぞれ τ_{pos}，τ_{neg} とする．本手法は，ニュースの拡散初期を想定し，一部のニュースのラベルを断定する STEP1 と，それ以降のニュースの拡散段階を想定し，時系列に沿ってニュースのラベルを断定する STEP2 に分けられる．

STEP1

STEP1 では，ニュースの発行直後（$t=t_0$）を想定するため，ニュースコンテンツのみを使用する．STEP2 で使用する BART を含めた複数の機械学習モデルを比較することで，STEP1 に最適な機械学習モデルを探索する．最適な機械学習モデルを用いて真偽の判別を行い，高い確率で真または偽と予測されたニュースについては，ラベルを断定する．ラベルが断定されなかったニュースは STEP2 でも真偽の判別を行う．一部のニュースについて初期段階で真偽を断定することで，フェイクニュースの早期検出を図る．

STEP2

STEP2 では，ニュースの拡散段階（$t=t_1,t_2,\ldots$）を想定する．t 時刻までの m 番目のソーシャルコンテクストを c_m とし，t 時刻までのソーシャルコンテクストを時系列順にソートしたものを，ソーシャルコンテクストの集合 $C_t=\{c_1,c_2,\ldots,c_m\}$ とする．また，t 時刻における入力を I_t，ニュースコンテンツを A とし，$I_t=\{A,C_t\}$ に対して各時刻において真偽を判別する．STEP1 と同様に，真偽の予測確率が閾値を超えたニュースはラベルを断定し，それ以降における判別を行わない．ソーシャルコンテクストの付与により，STEP1 よりも大きいシーケンス長を扱う必要があること，また Raza らの研究結果を考慮し，STEP2 では最大シーケンス長が 1024 の BART モデルを真偽の判別に用いる．

また，τ_{pos} =0.9，τ_{neg} =0.9 と定めた場合のラベルの断定フローを図 2 に示す．図 1 において，予測確率が書かれた長方形は，t 時刻におけるニュース記事を表す．灰色は予測ラベルが断定済みであることを，白色は予測ラベルが断定されていないことを示す．また，実線は該当の t 時刻に真偽の判別を行うことを，破線は判別を行わないことを示す．

4 データセット

本研究では，FakeNewsNet[1] [18] をデータセットとして使用した．FakeNewsNet はニュースコンテンツとソーシャルコンテクスト両方を含む英語のデータセットである．ニュースコンテンツには，ファクトチェックサイトである Politifact[2] と Gossipcop[3] のニュース記事が含まれる．Politifact は政治に関わるニュース記事のファクトチェックを主に提供し，Gossipcop は芸能に関わるニュース記事のファクトチェックを主に提供している．ソーシャルコンテクストには，それらのニュース記事を共有した Twitter[4] への投稿（以下ツイート）データが含まれる．我々は，FakeNewsNet データセットによって提供されたデータと TwitterAPI[5] を用いて，ニュース記事と関連するツイートデータを収集した．収集したニュース記事とツイートデータの統計情報を表 1 に示す．

1：https://github.com/KaiDMML/FakeNewsNet
2：https://www.politifact.com
3：https://www.gossipcop.com
4：https://www.twitter.com/
5：https://help.twitter.com/ja/rules-and-policies/twitter-api

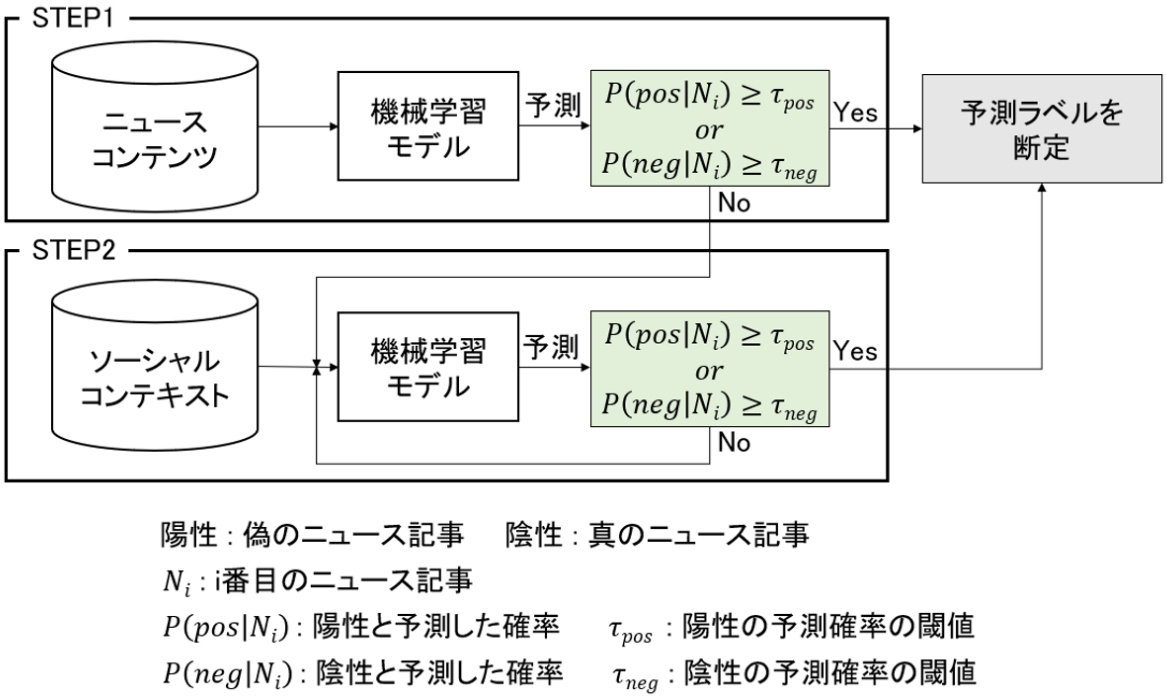

図 1　提案手法の全体像

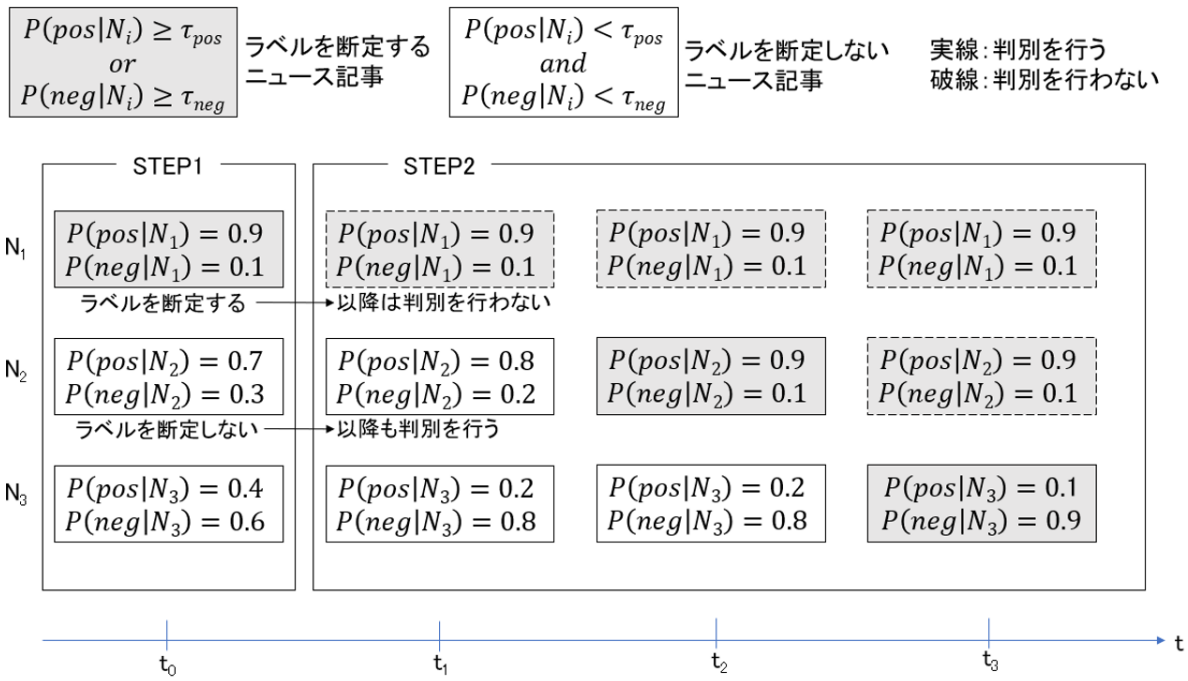

図 2　τ_{pos}=0.9，τ_{neg}=0.9 と定めた場合のラベルの断定フロー

表 1　収集したニュース記事とツイートデータの統計情報

データセット	Politifact		Gossipcop	
ラベル	真	偽	真	偽
ニュース記事（件）	356	360	13,837	4,447
ツイートデータ（件）	286,150	103,112	675,426	374,326

5　実験と検証

5.1　実験 1：ユーザーデータを用いた実験

ニュース記事の拡散に携わったユーザーに関する情報をフェイクニュースの検出に用いるため，実験 1 では，ニュース記事を共有するツイートを行ったユーザーのユーザーデータを，ソーシャルコンテクストとして使用した．

5.1.1　データセット

ダウンサンプリングの結果，実験 1 では 8,894 件の Gossipcop のニュース記事と，623,062 件のツイートデータを利用した．訓練データとテストデータの割合は 9:1 である．詳細を表 2 に示す．

表 2　実験 1：訓練データとテストデータの内訳

	真	偽	合計
訓練データ	4,002	4,002	8,004
テストデータ	445	445	890

5.1.2　ニュースコンテンツ

実験 1 では，ニュースコンテンツとして，ニュース記事のタイトルと本文を用いた．タイトルと本文を合算した，1 記事あたりの単語数を図 3 に示す．単語数の平均は 438 単語であり，8 割以上の記事が 500 単語以下で構成されていた．真偽どちらのラベルの記事においても，同様の分布が確認された．

5.1.3　ソーシャルコンテクスト

ニュース記事を Twitter 上に共有したツイートデータから，ツイートしたユーザーについて説明するテキスト（ユーザー自身が設定可能）を抽出し，ソーシャルコンテクストとして利用した．ニュース 1 記事あたりのツイートデータ数の平均は 70 件，1 ツイートユーザーあたりのユーザー説明欄の単語数の平均は 12 単語であった．また，図 4 に示すとおり，最初のツイートから 5 分以内の時間に最もツイートが集中していることが確認された．このことから，本研究の STEP2 では 5 分ごとにツイートのユーザーデータを付与し，真偽の判別を行う．

5.1.4　モデル

先行研究を参考に，STEP1 では STEP2 で用いる BART モデルの他に，古典的な機械学習モデルとしてロジスティック回帰 (LR) と非線形サポートベクタマシン (SVM)，深層学習モデルとして LSTM と textCNN を用いて実験を行った．また，BART モデルは facebook/bart-large-mnli[6]を利用した．設定したパラメーターは表 3 に示す．

6：https://huggingface.co/facebook/bart-large-mnli

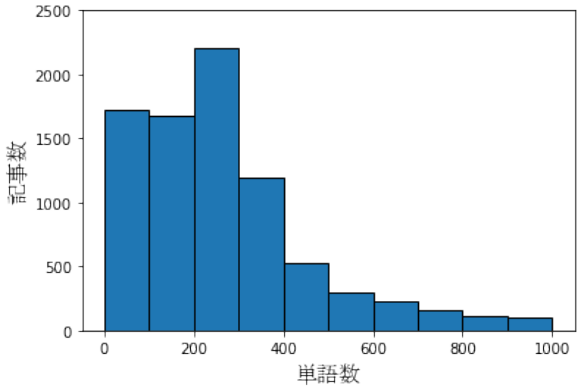

図 3　実験 1：1 記事あたりの単語数

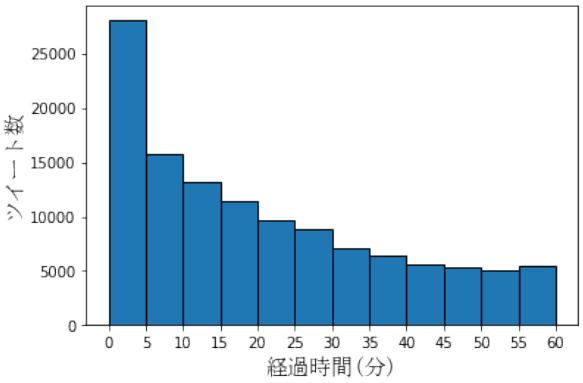

図 4　実験 1：5 分ごとのツイート数

表 3　BART モデルのパラメーター

最大シーケンス長	1024
学習率	1e-5
エポック数	2
バッチサイズ (学習時)	2
バッチサイズ (推論時)	64

5.1.5　評価指標

STEP1 では，ROC 曲線（Receiver Operatorating Characteristic curve，受信者動作特性曲線）と AUC（Area Under the ROC Curve，ROC 曲線の下の面積）スコアを用いてモデルの評価を行う．ROC 曲線は，検出閾値ごとに真陽性率（True Positive Rate：TPR）と偽陽性率（False Positive Rate：FPR）をプロットした曲線である．AUC スコアは ROC 曲線を積分した値であり，1 に近いほど良い検出モデルであることを表す．

STEP2 では，t を 5 分毎に設定し，精度と F1 値を求める．

5.1.6　実験結果

STEP1 におけるモデルごとの ROC 曲線を図 5 に，AUC スコアを図 4 に示す．BART モデルが最も高い AUC スコアを示していることから，低い偽陽性率で高い真陽性率を達成できるという観点において，BART モデルが最も優れたモデルであることが確認された．

STEP2 において，t を 5 分毎に設定した場合の精度と F1 値を表 6，7 に示す．ベースラインは，各時刻において全テストデータに対して真偽の判別を行った結果を用いて評価スコアを

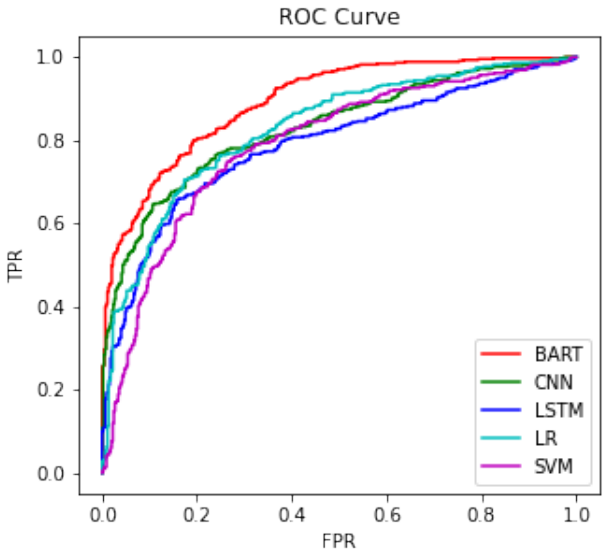

図 5　実験 1：STEP1 におけるモデルごとの ROC 曲線

表 4　実験 1：STEP1 におけるモデルごとの AUC スコア

モデル	AUC
LR	0.827
SVM	0.789
LSTM	0.788
textCNN	0.830
BART	**0.893**

算出している．$\tau_{pos} = 0.9$，$\tau_{neg} = 0.9$ という閾値を定めてラベルを断定した場合，ベースラインと比較して，最初のツイートから 5 分後と 10 分後の時刻における F1 値が向上することが確認された．また，テストデータ 890 件に対する，ラベル断定済みデータの割合を図 8 に示す．閾値を $\tau_{pos} = 0.9$，$\tau_{neg} = 0.9$ とした場合，5 分後，10 分後の時刻において，9 割近いデータがラベル断定済みであることが確認された．これらのことから，提案手法がフェイクニュースの早期検出に貢献することが確認された．

5.2　実験 2：リプライツイートデータを用いた実験

実験 2 では，ニュースの拡散に携わるユーザーの行動を直接的に特徴量として反映させるため，ツイートのテキストをソーシャルコンテクストとして利用した．ニュース記事を共有したツイートは Twitter の仕様上同じテキストになるため，共有ツイートに対する返信（リプライ）ツイートのテキストを用いた．

5.2.1　データセット

実験 2 では，なるべく多くのリプライツイートを利用するため，Politifact と Gossipcop のニュース記事 9,612 件と，関連する 132,893 件のツイートデータを利用した．実験 2 で使用したデータセットの詳細を表 5 に示す．訓練データとテストデータの割合は 8:2 である．ただし，ニュース記事に対するリプライツイートの総数に偏りがあるため，データセットをリプライツイート数によって分類した後，8:2 の割合で訓練データとテストデータに分割を行った．また，リプライツイートが存在し

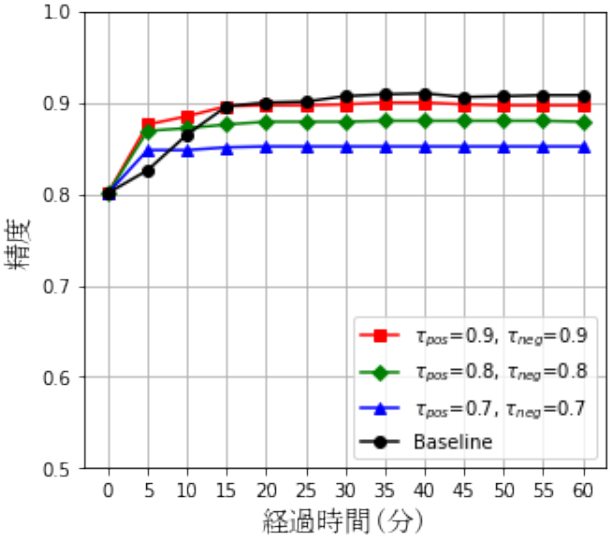

図 6　実験 1：STEP2 における 5 分ごとの精度

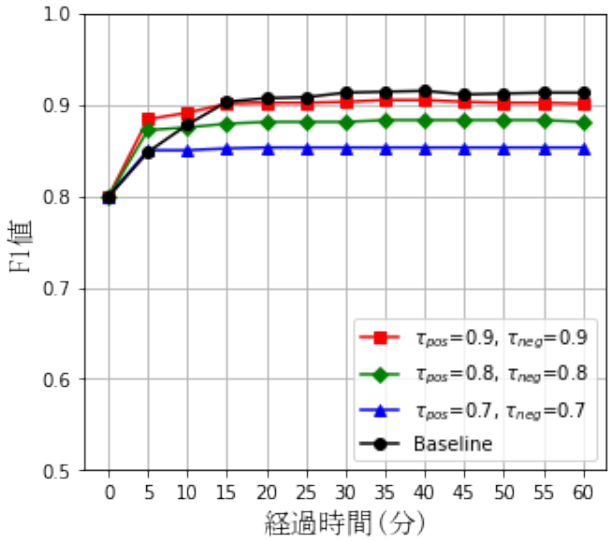

図 7　実験 1：STEP2 における 5 分ごとの F1 値

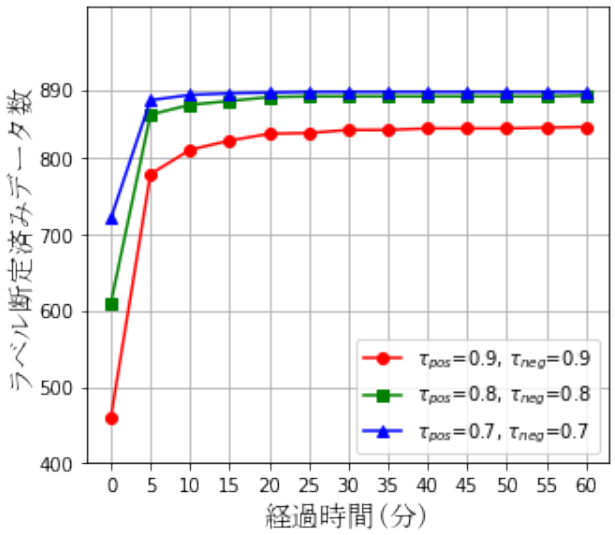

図 8　実験 1：STEP2 における 5 分ごとのラベル断定済みデータ数

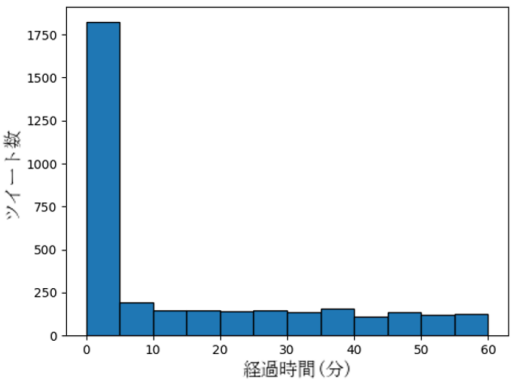

図 9　実験 2：5 分ごとのツイート数

ないニュース記事にはソーシャルコンテクストの付与ができないため，STEP2 ではリプライツイートが存在するニュース記事のみを訓練データとして用いている．同様に，STEP2 の推論時には，リプライツイートが存在するニュース記事についてのみ真偽の判別を行い，リプライツイートが存在しないニュース記事は STEP1 の予測ラベルを用いることで評価を行う．

表 5　実験 2：訓練データとテストデータの内訳

データセット	Politifact		Gossipcop		
ラベル	真	偽	真	偽	合計
訓練データ (STEP1 リプライ有)	172	109	259	682	1,222
訓練データ (STEP1 リプライ無)	116	175	3,298	2,874	7,685
訓練データ (STEP2)	172	109	259	682	1,222
テストデータ	74	72	890	891	1,927
テストデータ (リプライ有)	44	28	66	171	309

5.2.2　ニュースコンテンツ

実験 1 と同様に，実験 2 では，ニュース記事のタイトルと本文をニュースコンテンツとして用いた．

5.2.3　ソーシャルコンテクスト

ニュース記事を Twitter 上に共有したツイートに対するリプライツイートのテキストを，ソーシャルコンテクストとして使用した．ニュース 1 記事あたりのリプライツイート数の中央値は 4 件，リプライツイート 1 件あたりの単語数の平均は 20 単語であった．また，図 9 に示すとおり，最初のツイートから 5 分以内の時間に最もツイートが集中していることが確認された．しかし，ユーザーデータと比べてリプライツイートの件数が非常に少なく，ツイート時刻のばらつきも大きいことから，実験 2 では，時刻 t_{n-1} から t_n の間に 5 件のリプライツイートが発生すると仮定し，時系列順に 5 件ずつリプライツイートデータを入力に付与することで，実験を行った．

5.2.4　モ　デ　ル

実験 1 の結果から，実験 1 と同様の BART モデルを STEP1 における判別に使用した．設定したパラメーターも実験 1 と同様である．

5.2.5　評 価 指 標

STEP1，STEP2 を通して，時系列順に 5 件ずつリプライツイートデータを入力に付与し，精度と F1 値を求める．

5.2.6　実 験 結 果

実験 2 における精度と F1 値を図 10，11 に示す．実験 1 と同様に，ベースラインは各時刻において全テストデータに対して真偽の判別を行った結果を用いて，評価スコアを算出している．リプライツイートを付与することによって精度と F1 値が向上したベースラインに対し，提案手法は F1 値のみが向上し，精度の向上は見られなかった．また，いずれの評価スコアにおいてもベースラインのスコアが高い値を示している．

5.2.7　実験結果の分析

実験 2 について，テストデータのうち，リプライツイートが存在する 309 件について分析を行う．上記のデータのうち，陽性の予測確率が τ_{pos} を超えた場合のみ予測ラベルを陽性とし，τ_{pos} を下回った場合には予測ラベルを陰性とした場合の適合率を図 12 に示す．同様の場合の再現率を図 13 に，F1 値を図 14 に示す．リプライツイートを特徴量として付与することにより，高い適合率を保ったまま再現率と F1 値が向上していることから，リプライツイートがフェイクニュースの検出に役立っていることが確認される．また，予測ラベルを断定する閾値が高いほど，再現率・F1 値・ラベル断定済みデータ数の増加の収束が遅いことから，閾値が高いほどより多くのリプライツイートの情報を考慮した判別を行う可能性が示唆される．

また，t_0，t_1 における陽性の予測確率の分布を，それぞれ図 16，図 17 に示す．ここで，fake は偽のニュース記事（正解ラベルが陽性）の予測確率を，true は真のニュース記事（正解ラベルが陰性）の予測確率を示す．図 16，図 17 から，リプライツイートを付与することにより，偽のニュース記事に対して陽性の予測確率が上がっていることが確認される．一方，真のニュース記事に対しても陽性の予測確率が上がっていることが確認されるため，偽陽性の数が増加することが，提案手法の実験 2 における精度や F1 値が向上しない原因になっていると考えられる．t_1 以降は，予測確率の分布に大きな変化は見られなかった．

詳細を分析した結果，t_0 において陽性の予測確率が 0.3 未満であった偽のニュース記事が，t_1 において 0.9 以上の確率で陽性と予測される場合が散見された．このことから，t_n における予測確率に加え，t_n における予測確率と t_{n+1} における予測確率の変化量を考慮した予測ラベルの断定条件を加えることにより，リプライツイートを用いた提案手法の検出精度の向上が可能であると考えられる．

6　まとめと今後の課題

本研究では，ニュースコンテンツと関連するソーシャルコンテクストを用いた，機械学習による 2STEP のフェイクニュース早期検出手法を提案した．ニュースの発行直後を想定した STEP1 では，BART モデルが最も高い精度を示し，予測確率の信憑性にも優れることが確認された．ニュースの拡散段階を想定した STEP2 では，ユーザーデータをソーシャルコンテクストとして用いた場合に，予測確率 0.9 以上のデータの真偽のラベルを断定することにより，ニュースの拡散初期段階におい

て 9 割近くのデータのラベルを断定したうえで，ベースラインよりも高い検出精度を実現できることが確認された．このことから，ユーザーデータを用いた提案手法がフェイクニュースの早期検出に貢献することが確認された．

今後は，予測確率の変化量を考慮した条件を予測ラベルの断定条件に追加することにより，ユーザーの行動を直接的に反映したリプライツイートなどの情報を用いた場合の検出精度の向上を目指す．また，BART を含む各モデルのパラメーターや，STEP2 におけるニュースコンテンツとソーシャルコンテクストのシーケンス長の割合を変更することによる，検出の精度と早期検出への影響を調査する．

謝　　辞

本研究は JSPS 科研費 JP22K12271，JP19K12230 の助成を受けたものである．

文　　献

[1] Kai Shu, Amy Sliva, Suhang Wang, Jiliang Tang and Huan Liu. Fake News Detection on Social Media: A Data Mining Perspective. *KDD*, 2017.
[2] Soroush Vosoughi, Deb Roy and Sinan Aral. The spread of true and false news online. *Science*, 2018.
[3] Reza Zafarani, Xinyi Zhou, Kai Shu and Huan Liu. Fake News Research: Theories, Detection Strategies, and Open Problems. *KDD*, 2019.
[4] Xinyi Zhou and Reza Zafarani. A survey of fake news: Fundamental theories, detection methods, and opportunities. *ACM Comput. Surv.*, 2020.
[5] Konstantin Pogorelov, Daniel Thilo Schroeder, Stefan Brenner, and Johannes Langguth. Fakenews: Corona virus and conspiracies multimedia analysis task at mediaeval 2021. *MediaEval*, 2021.
[6] Jiawei Zhang, Bowen Dong and Philip S. Yu. FAKEDETECTOR: Effective Fake News Detection with Deep Diffusive Neural Network. *ICDE*, 2020.
[7] Changhe Song, Cunchao Tu, Cheng Yang, Zhiyuan Liu, Maosong Sun. CED: Credible Early Detection of Social Media Rumors. *TKDE*, 2020.
[8] Xinyi Zhou, Jindi Wu and Reza Zafarani. SAFE: Similarity-Aware Multi-Modal Fake News Detection. *PAKDD*, 2020.
[9] Xinyi Zhou, Reza Zafarani, Emilio Ferrara. From Fake News to #FakeNews: Mining Direct and Indirect Relationships among Hashtags for Fake News Detection. arXiv preprint, 2022.
[10] Xinyi Zhou, Atishay Jain, Vir V. Phoha and Reza Zafarani. Fake News Early Detection: A Theory-driven Model. *ACM Digital Threats: Res. Pract.*, 2019.
[11] Xinyi Zhou, Reza Zafarani. Network-based Fake News Detection: A Pattern-driven Approach. *KDD*, 2019.
[12] Shaina Raza and Chen Ding. Fake news detection based on news content and social contexts: a transformer-based approach. *Int. J. Data. Sci. Anal.*, 2022.
[13] 山中仁斗， 張建偉. 発生規模と時系列を考慮した twitter イベントにおける偽情報の早期自動検出. 第 12 回データ工学と情報マネジメントに関するフォーラム（*DEIM 2020*）, 2020.
[14] Mike Lewis, Yinhan Liu, Naman Goyal, Marjan Ghazvininejad, Abdelrahman Mohamed, Omer Levy, Ves Stoyanov, and Luke Zettlemoyer. Bart: Denoising sequence-to-sequence pre-training for natural language generation, translation, and comprehension. *ACL*, 2020.
[15] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is all you need. *NeurIPS*, 2017.
[16] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. BERT: Pre-training of deep bidirectional transformers for language understanding. *NAACL-HCT*, 2019.
[17] Alec Radford, Karthik Narasimhan, Tim Salimans, Ilya Sutskever, et al. Improving language understanding by generative pre-training. OpenAI Blog, 2018.
[18] Kai Shu, Deepak Mahudeswaran, Suhang Wang, Dongwon Lee, and Huan Liu. Fakenewsnet: A data repository with news content, social context and spatialtemporal information for studying fake news on social media. *Big Data*, 2018.
[19] Maurício Gruppi, Benjamin D. Horne and Sibel Adalı. NELA-GT-2019: A large multi-labelled news dataset for the study of misinformation in news articles. arXiv preprint, 2020.
[20] Kai Nakamura, Sharon Levy, and William Yang Wang. r/fakeddit: a new multi-modal benchmark dataset for fine-grained fake news detection. *LREC*, 2019.

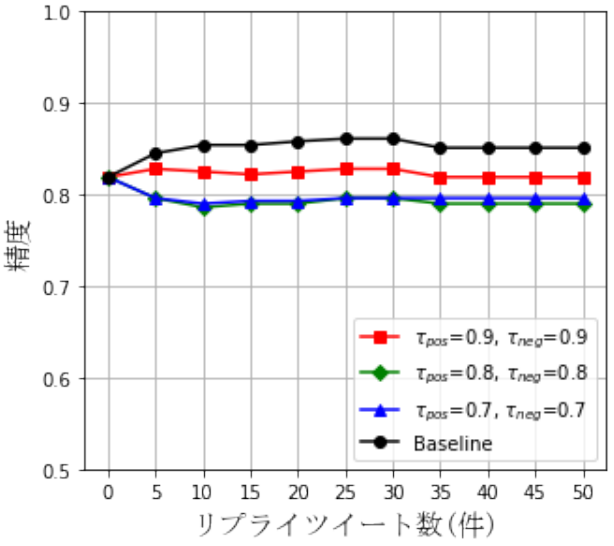

図 10　実験 2：リプライツイート 5 件ごとの精度

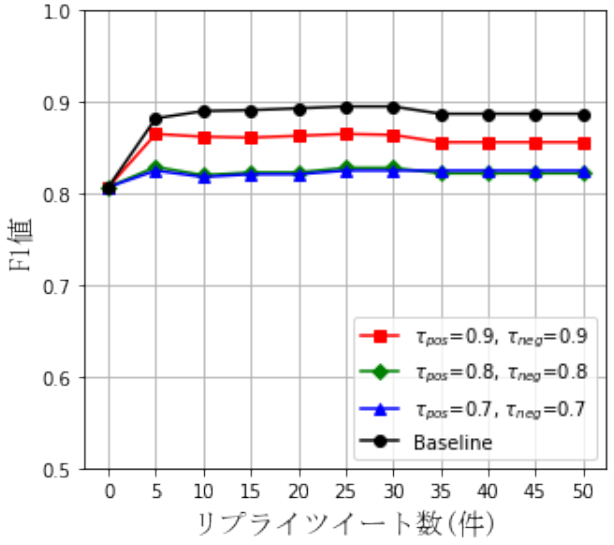

図 11　実験 2：リプライツイート 5 件ごとの F1 値

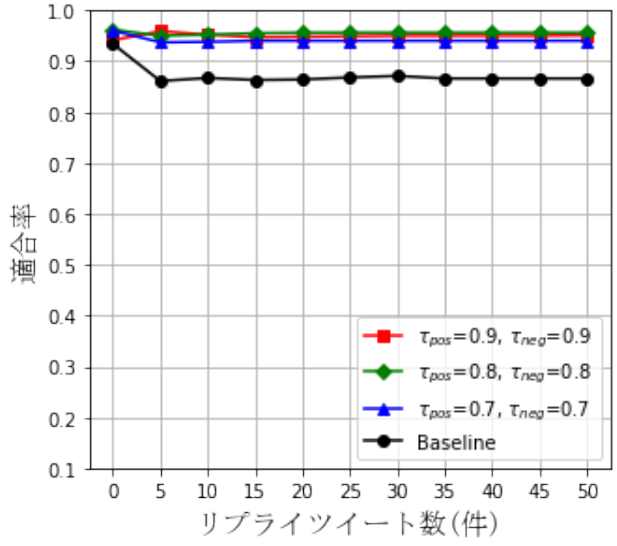

図 12　実験 2：陽性の予測確率が閾値以上のデータに対する適合率

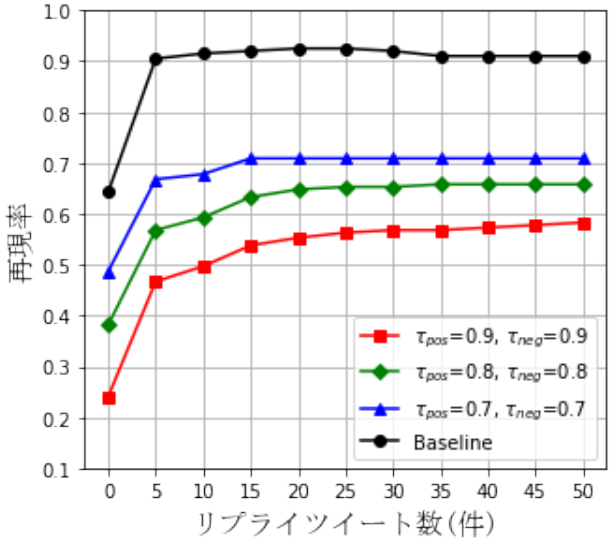

図 13　実験 2：陽性の予測確率が閾値以上のデータに対する再現率

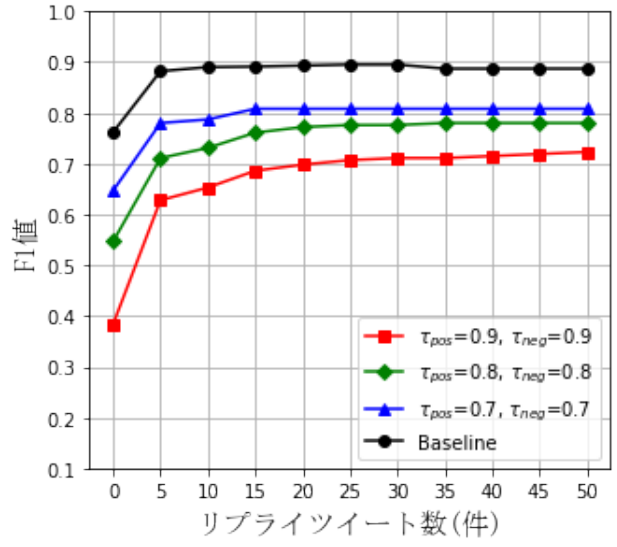

図 14　実験 2：陽性の予測確率が閾値以上のデータに対する F1 値

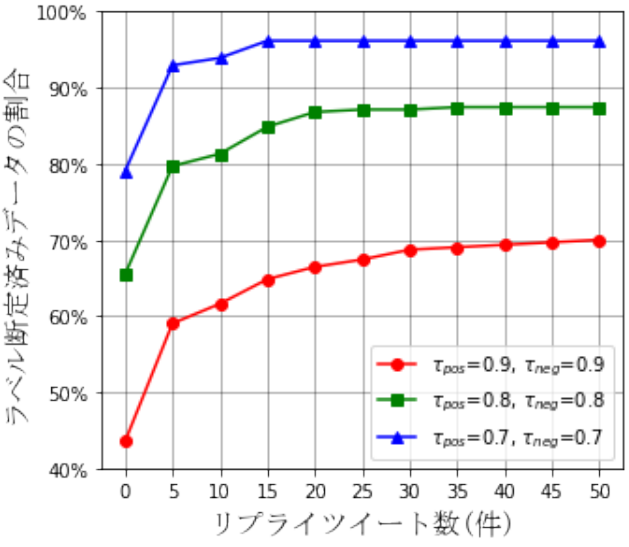

図 15　実験 2：閾値以上の予測確率を示したデータの割合

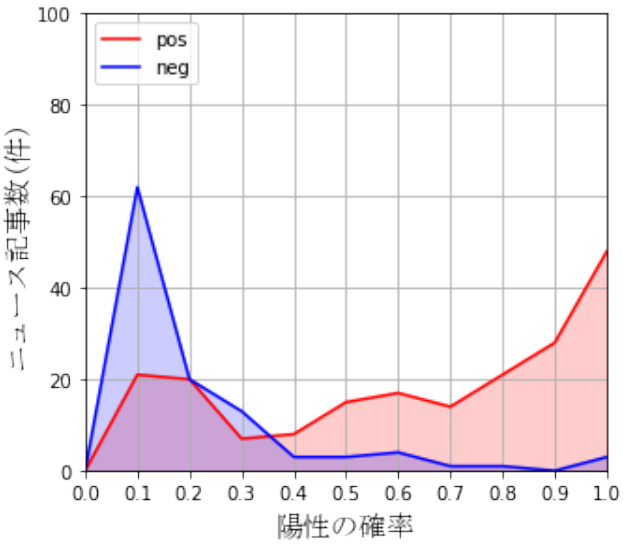

図 16　実験 2：t_0 における陽性の予測確率の分布

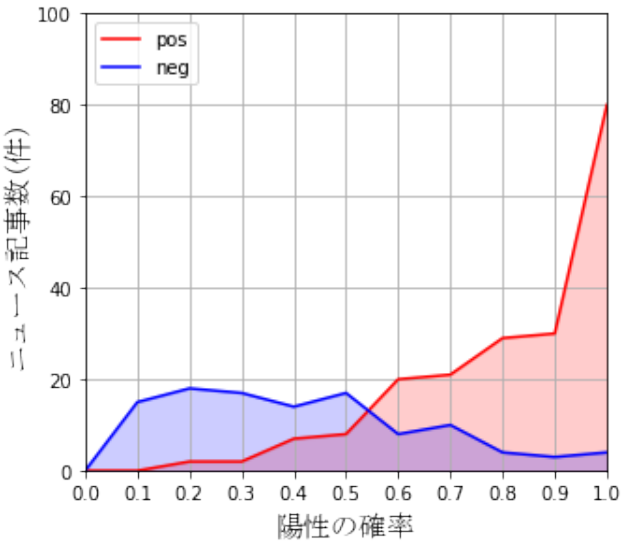

図 17　実験 2：t_1 における陽性の予測確率の分布