A-033

信頼性評価ネットワークを用いたフェイクニュースに関与するユーザーの特徴抽出モデル
Feature Extraction Model of Users Engaged in Fake News Using Credibility Rating Network

星 沙耶香‡　　穴田 一‡
Sayaka Hoshi　Hajime Anada

1. はじめに

ソーシャルメディアの影響度が増している昨今，フェイクニュースを自動で検出する研究が大きな注目を集めている。フェイクニュースという言葉は、広義の定義[1]と狭義の定義[1]が存在し、広義の定義では、情報の信憑性のみを強調して情報の意図を考慮しないのに対して、狭義の定義では意図的に発信された虚偽の情報としている。

村山ら[1]によると、多くの研究が発信者の意図を考慮した狭義のフェイクニュースを採用しているにもかかわらず、使用するデータセット構築は、事実確認サイトの判断に基づくニュースの事実性にのみ着目し、意図には着目していない。そのため、既存のデータセットに基づいて構築され検出モデルの多くは、ソーシャルメディアに投稿された情報が真実か嘘かを判断することに焦点を当てている。一方で、昨今のフェイクニュースの社会への影響から、意図的に読者を誤解させたり、だましたり、誘ったりする狭義のフェイクニュースを検出することも重要な課題である。

そこで本研究では、狭義のフェイクニュース検出を補助するシステムの構築を目指し、石田ら[2]が提案した信用度評価モデルを改良して、ニュースの発信者及び反応したユーザーの信用度を動的に評価するモデルを提案し、ユーザーの評価値の妥当性を検証する。

2. 既存研究

2.1 動的関係ネットワーク

石田らは、調べたいニュースと関連する他のニュースの内容の事実関係を相対的整合性により動的に決める、動的関係ネットワーク(Dynamic relational Network)を用いた信用度評価モデルを提案した。動的関係ネットワークは SNS 上の人間関係のように、時間が経過すると構造が変化するネットワークを指し、ネットワーク上のノードは互いの持つ値を信用可能であるか評価する。評価はノード間の有向エッジによって行う。

このエッジには、二つのノード i, j が有する値の関係によって正または負の値を設定し、正の場合は整合、負の場合は矛盾を意味する。一方、ネットワーク上のノード i, j は実数の信用度$R_i(t) \in [0,1]$を持つ。信用度はノードの有する値がどの程度信用可能かを表し、1 に近いほど信用でき、ネットワーク上ではノードが活性化する。以下にノードの評価式を示す。

$$\frac{dr_i(t)}{dt} = \sum_j (T_{ij} + T_{ji} - 1)R_j(t) - r_i(t)$$
$$R_i(t) = \frac{1}{1 + exp(-r_i(t))} \quad (1)$$

r_i: 正規化前の信用度
R_i: r_iを正規化した値
T_{ij}:ノード i のノード j に対する評価値で+1 か-1 をとる
　ノード i, j 間に評価が存在しない場合は 0 をとる

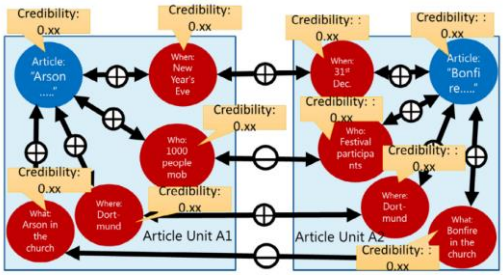

図 1 動的関係ネットワーク生成例

2.2 信用度評価モデル

動的関係ネットワークは、結論ノードと事実ノードから構成される。結論ノードは、事実情報の評価値によって最終的に信用度を評価するノードである。また、事実ノードは、5W1H の情報をラベルとして持つ。

まず、評価したい記事の信用度を算出するための結論ノードを配置する。次に、評価したい記事の事実 (5W1H に対応する内容)をラベルとした、事実ノードを配置する。その後、結論ノードと事実ノードの間に双方向のエッジを生成する。続いて、比較対象である関連記事も同様にノードを配置する。さらに、同じ種類のラベルを持つ事実ノード間にも双方向のエッジを生成し、最終的に時間経過によって収束した各ノードの信用度を結果とする。図 1 に生成したネットワークの例を示す。

2.2.1 結論ノード

結論ノードは、事実情報の評価値によって最終的に信用度を評価するノードである。このノードと事実ノード間のエッジの重みは+1 である。ノードを設置する段階でのニュースは全て正しいとして、ラベルは「『述べられている内容は真実である』」を持つ。初期信用度は最大である 1 を与える。

2.2.2 事実ノード

事実ノードはニュースで事実と述べられている内容(5W1H)をラベルとして持ち、そのラベルの信用度を評価するノードである。事実ノード間のエッジは、ラベルの内容が同じ場合は+1、内容が異なる場合は-1 で重みづけされる。ノードを設置する段階でのニュースは全て正しいとして、初期信用度は最大である 1 を与える。

3. 提案手法

Yi-Ju[3]らは、広義のフェイクニュース検出の結果、不審なユーザーは、アカウントが認証されていない、アカウント作成時間が短い、ユーザーの説明文が短い、などの特徴があることを示し、ニュース発信者の意図を理解する上で有益だと考察している。また、悪意を持って作成されたニュースは、そうでないものに比べて説得力が増すことを目的としており、悪意のあるユーザーはソーシャルメディ

ア上での知名度を上げるために偽ニュースの伝播に参加することが一般的に知られている[1]。

したがって、ニュース発信者の意図を捉えるには、発信者及び反応したユーザーの疑わしさを評価し、ニュースの伝搬をユーザーレベルで捉える必要がある。そこで本研究では、発信者及び反応したユーザーの疑わしさを石田らのモデルを用いて決定し、ニュースが拡散されるにつれて値がどのように変化するのか検証する。以降、石田らの評価モデルを改良した点を述べる。

3.1 信用度評価モデル

あるニュース S が初めて投稿されてから受け取る一連のユーザー反応を、一定時間間隔$\Delta t_n (n = 0, \ldots, N)$ に分けて考える。提案手法では、Δt_nで初めてニュースに反応したユーザーUと、既に反応したユーザーU'に関して動的関係ネットワークを作成し、Δt_nにおける互いの信用度を算出する。ユーザーU'は、ユーザーUの一連の全ての親ノードを意味する。例えば、図 2 のようにニュースが拡散した場合，Δt_0 では$U = \{A, B\}$, $U' = \{S\}$ となり，Δt_1 では$U = \{C, D, E\}, U' = \{B, S\}$となる。

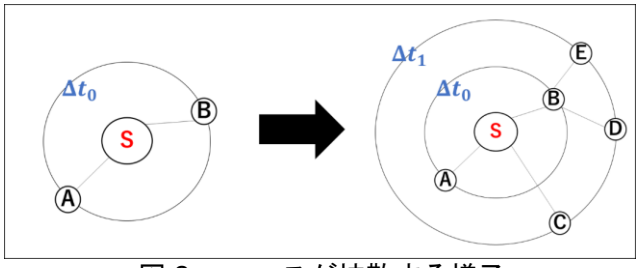

図 2 ニュースが拡散する様子

動的関係ネットワークの作成手順を以下に示す。

I. ユーザーUに関して、信用度を評価するための結論ノードを配置する。次に、評価したいユーザーの事実をラベルとした、事実ノードを配置する。事実ノードは、新たに設定したユーザーのツイートにのみ依存する情報をラベルとして持つ(詳細は 3.1.1 で述べる)。その後、結論ノードと事実ノードの間と、同じ種類のラベルを持つ事実ノード間に双方向のエッジを生成する。

II. ユーザーU'に関して，信用度を評価するための結論ノードを配置する。その後ユーザーUの結論ノードと合わせて、拡散ネットワークの構造上、繋がっているユーザー同士の結論ノード間に、双方向のエッジを生成する。最終的に時間経過によって収束した各ノードの信用度をΔt_n における結果とする。

ここで、ユーザーUの結論ノードと事実ノードの初期信用度は 1 、ユーザーU'の結論ノードの初期信用度はその時点までに決定した値を用いる。手順 I , II を n=N まで繰り返すことで信用度を更新する。

例として、事実ノードが 2 種類のラベル(f1, f2)を持つとした時、図 2 で示したΔt_1における動的関係ネットワークを生成する様子を図 3 に示す。

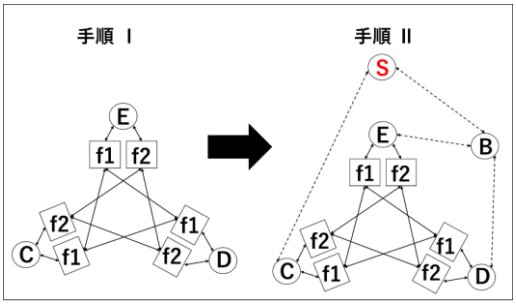

図 3 Δt_1での動的関係ネットワーク生成

図形は丸が結論ノード、四角形が事実ノードを表す。手順 I を実線, 手順 II を破線で表す。

3.1.1 事実ノード

石田らのモデルでは、事実ノードはニュースの 5W1H に対応した情報をラベルに持つが、twitter の会話情報における投稿はテキスト文が短く、同様にラベル付けするのは困難である。そこで、提案手法の事実ノードは、[4]を参考にして、ニュースにのみ依存する以下の 4 種類の特徴をラベルに持つとする。

①語彙に関する特徴

テキストのベクトル表現(word2vec でツイート内の単語をベクトル化して平均した値)、POS タグ(各品詞の出現回数)、否定語使用の有無

②ツイート内容の書式に関する特徴

ツイートの長さ、大文字の使用割合、単語数

③句読点に関する特徴

「?」の使用有無、「!」の使用有無、「.」の使用有無

④ツイート自体の書式に関する特徴

URL の有無、画像の有無

特徴①～④のうち、①のテキストのベクトル表現と POS タグ、②は数値データである。そこで、①のテキストのベクトル表現」は、比較するノード同士でコサイン類似度を算出し、0.6 以上の場合は+1, 0.6 未満の場合は-1 でエッジを重み付ける。さらに、①の POS タグはジャカード係数を算出し、同様に 0.6 以上の場合は+1、0.6 未満の場合は-1 でエッジを重み付ける。②に関しては、この順に 3 つ並べてベクトル化したものを 1 つのノードとして、比較するノード同士でコサイン類似度を算出し、0.6 以上の場合は+1, 0.6 未満の場合は-1 でエッジを重み付ける。閾値の設定は[5]を参考にした。それ以外の特徴は 2 値データのため、事実ノード間のエッジは、ラベルの内容が同じ場合は+1、内容が異なる場合は-1 で重みづけされる。まとめると、提案手法の事実ノードは、各ユーザーにつき 8 個となる。

4. 評価実験

実験には、9 種類のニュース速報に関連する噂の会話を Twitter から収集した PHEME データセット[6]を用いる。このデータセットに含まれる噂、非噂にラベル付けされた各投稿について、提案手法を用いて動的関係ネットワークを生成し、信用度を得られるか投稿の種類ごとに比較をしながら評価を行う。

結果、考察は発表時に述べる。

参考文献

[1] Murayama Taichi, Hisada, Shohei, Uehara, Makoto, Wakamiya, Shoko, Aramaki, Eiji, “Annotation-Scheme Reconstruction for "Fake News" and Japanese Fake News Dataset”, (2022).
[2] Ishida Yoshiteru, Kuraya Sanae, “Fake News and its Credibility Evaluation by Dynamic Relational Networks: A Bottom up Approach”, Procedia Computer Science, Vol.126 (2018).
[3] Yi-Ju Lu, Cheng-Te Li, “GCAN: Graph-aware Co-Attention Networks for Explainable Fake News Detection on Social Media”, ACL (2020)
[4] Zubiaga Arkaitz, Liakata Maria, Procter Rob, “Exploiting Context for Rumour Detection in Social Media”, LNISA, Vol.10539 (2017)
[5] Zhao, Zhe, Resnick, Paul, Mei, Qiaozhu, “Enquiring minds: Early detection of rumors in social media from enquiry posts”, WWW (2015)
[6] Elena Kochkina, Maria Liakata, Arkaitz Zubiaga, https://figshare.com/articles/dataset/PHEME_dataset_for_Rumour_Detection_and_Veracity_Classification/6392078?file=11767817