Vol.2020-DPS-182 No.61
Vol.2020-CSEC-88 No.61
2020/3/13

個人間の信頼と不信頼を動的に更新するフェイクニュースの伝搬モデルの提案

由川 拳都[1,a)]　阿波 拓海[2,b)]　草野 理沙[2,c)]　市野 将嗣[1,d)]　吉浦 裕[1,e)]

概要：昨今，フェイクニュースが社会的な問題となっている．代表的な対策として，マスメディアによるファクトチェックと機械学習などを用いたフェイクニュースの検知がある．しかし，フェイクニュースを信じる人は，事実を指摘しても受け入れず，逆に自分の信念を強める傾向があるので，これらの対策は有効とは限らない．そこで，フェイクニュースの性質を明らかにするための伝搬モデルの研究が重要であると考える．意見の伝搬をモデル化した先行研究として，他人の意見への信用度から自分の信念に対してベイズ更新を行うモデルがあるが，信用度が一律かつ固定であり，現実的ではない．また，ベイズ更新の式に誤りがある．そこで，本稿では，他人ごとに信用度を設定し，かつ複数の他人の意見から自分の信念に対してベイズ更新を行うとともに，EM アルゴリズムにより信用度を動的に更新する手法を提案する．また，この信用度と信念の更新に基づいて，ネットワーク全体の信念および意見の伝搬をモデル化する．シミュレーションによる評価実験の結果，ネットワークの形態に関わらず意見が決まりやすくなり，かつ正しい意見が拡散されやすくなることが明らかとなった．

キーワード：フェイクニュース，ベイズの定理，EM アルゴリズム，マルチエージェントモデル

1. はじめに

情報化社会が進むにつれ，人々が情報を伝え，それを受け取る手段が変化している．従来では，新聞やテレビなどのマスメディアが独占的にその役割を担っていた．しかし，現在ではブログやソーシャルネットワーキングサービス (以下，SNS と呼ぶ) といったソーシャルメディアがその役割を担っている．そのため，個人が簡単に情報発信できるようになり，有益な情報が流れやすくなった一方，デマや嘘といった不確かな情報も流れやすくなった．これらがフェイクニュースと呼ばれるようになり，2016 年の米大統領選以降，注目を集めるとともに社会的な問題となっている．

このフェイクニュースによって国内外で様々な事件が発生している．例として，米国では「ワシントンのあるピザ店が小児性愛者集団の拠点となっておりヒラリー・クリントンがその集団に関与している」という虚偽のニュースを見た男性がそのピザ屋で発砲し，逮捕される事件が発生した．また，国内では 2016 年の熊本事件直後に「熊本市動物園からライオンが逃げた」という旨の虚偽の投稿が SNS 上で拡散され，その投稿を行った男性が逮捕されるという事件が発生した．現在のフェイクニュース対策の一つとして，マスメディアが連携し，事実確認を行うファクトチェックと呼ばれるものがある．ファクトチェックは人手で行うため，時間と費用を要する．また， 人間の心理現象として，事実を指摘しても受け入れず，逆に自分が信じたい意見への確信を強めるようになるバックファイヤー効果と呼ばれるものがある [1]．この性質は，ソーシャルメディアの台頭とともに顕著になっている [2]．以上より，ファクトチェックが効果的な対策とは言い難い．

フェイクニュースに関する代表的な先行研究としてフェイクニュースの検知がある [3] [4]．しかし，この研究は検知をする側とされる側のいたちごっこになる傾向にある．また，バックファイヤー効果から真実を無視する場合もある．よって，フェイクニュースの検知は効果的な対策とは言い難い．したがって，効果的なフェイクニュース対策を生み出すには，フェイクニュースの性質やその意見伝搬を明らかにする必要があると考える．合理的な人間の意見

[1] 電気通信大学情報理工学研究科
Gradudate School of Informatics and Engineering, The University of Electro-Communications
[2] 電気通信大学情報理工学域
School of Informatics and Engineering, The University of Electro-Communications
[a)] k-yoshikawa@uec.ac.jp
[b)] taku.awa@uec.ac.jp
[c)] r1.Kusano@uec.ac.jp
[d)] ichino@inf.uec.ac.jp
[e)] yoshiura@uec.ac.jp

伝搬をモデル化している先行研究として，Pryymak らは AAT(Autonomous Adaptive Tuning) と呼ばれる手法を提案した [5]．そして，AAT により誤った意見の存在下でも真実と同じ意見を共有できることを示した．しかし，AAT では以下に述べる問題がある．はじめに，AAT では他人の意見を等しく信用し，他人の意見の信用度も一定であることを前提としている．しかし，この前提は実際の人間の振る舞いを反映しているとは言い難い．次に，AAT では他人の意見を多少なりとも信用することを前提としている．しかし，[1] [2] で報告されているバックファイヤー効果を起こす人が一定数存在することを考慮すると，必ずしも妥当な前提とは限らない．よって，AAT はフェイクニュースの伝搬を表現できる手法ではない．また，Pryymak らは，AAT を適用するために必要な意見の更新式をベイズの定理に基づいて定義しているが，この更新式には誤りがある．

そこで，本稿では先行研究 [5] の問題点を解決することで，フェイクニュースにおける意見伝搬を考慮したマルチエージェントモデルを提案する．そして，このモデルを通してフェイクニュースの正確な分析を行うことを研究目的とする．

2. 先行研究

2.1 概要

代表的な研究としてフェイクニュースの検知がある．例として，ニュースサイトの文章特徴を機械学習により学習することで，ニュースサイトの真偽を判別した研究がある [3]．また，ギブスサンプリングに基づいたアルゴリズムにより，SNS 上のフェイクニュースに対して言及しているユーザーの信頼性を推定することで，SNS 上のフェイクニュースを検知した研究もある [4]．

フェイクニュースの研究には攻撃の研究も存在する [6]．この研究では，深層学習の手法の一つであるリカレントニューラルネットワークを用いて虚偽レビューを生成した．そして，そのレビューが機械学習と人間に虚偽と見破られるかどうか検証した．

また，SNS の一つである Facebook 上で，フェイクニュースに関与するユーザの心理を分析した研究もある [2]．この研究により，偏った意見を持つユーザとその友人が同じ行動をとることで，意見の偏ったコミュニティが形成されることを明らかにした．また，偏った意見を持つ人物には，1 章で述べたバックファイヤー効果が起こることも示した．

そして，フェイクニュースの拡散について着目した研究もある [7] [8]．この研究により，SNS 上で目撃されるフェイクニュースは真実よりも広範囲に早く伝わりやすいという性質があることが判明し， 特に政治に関するフェイクニュースにおいてその性質が顕著に表れることが示された．そして，フェイクニュースの拡散には人間だけではなく，特定のキーワードに対する投稿や共有を自動で行うプログラムである「ボット」が関与していることも示された．また，情報拡散をモデル化した研究として，SNS 上で観測された実際のデマが拡散され，訂正される様子を感染症モデルによりモデル化した研究もある [9]．

2.2 マルチエージェントモデルによる意見共有

Glinton らは，他人の意見に誤りが存在する状況で，集団が正しい意見の共有を目指す問題を意見共有問題と呼び，これを定式化した [10]．ここで，正しい意見とは True もしくは False で定義される外部の客観的事実が存在することを前提としたときに，その事実と同じ真偽を持っている状態であることを指す．

意見共有問題において，Pryymak らは大規模な集団間で正しい意見を共有するためのアルゴリズムとして，AAT(Autonomous Adaptive Tuning) を提案した [5]．AAT では，個人が信用度というパラメータを持つ．このパラメータは，他人の意見の信じやすさに相当し，$[0.5, 1.0]$ の範囲で定義されている．この信用度を，複数回のシミュレーションを通して調整することで，集団として正しい意見を共有できるようにする．その結果，集団のネットワークの形態に関わらず正しい意見を共有できることを示した．また，AAT では，ベイズの定理に基づいて意見の更新式を定義し，その更新式により意見を決定している．そのため，人間が合理的に意見を形成する様子をモデル化している．しかし，1 回のシミュレーションでは他人の意見への信用度が一定かつ他人に応じて信用度が区別されていないという問題がある．そして，信用度が 0.5 以上であるため，他人の意見を多少なりとも信用することを前提としているが，バックファイヤー効果を考慮するとこ必ずしも妥当な前提とは限らない．よって，フェイクニュースの伝搬を表現できない．さらに，ベイズの定理に基づいた意見の更新式には 3.1 節 (iii) 項で述べる誤りがある．

2.3 EM アルゴリズムによる他人の信用度の推定

Wang らは，情報の報告者の信用度が不確かな状態で，最も信頼できる報告者とその事柄を見つける方法を提案した [11]．彼らは，報告者が報告した複数の情報に関する真偽に対して，最尤推定法の一つである EM アルゴリズムを適用することで，報告者の信用度を推定した．この結果，真偽が相反する報告を受け取った場合や報告の数が少ない場合でも最も信頼できる報告者とその事柄を見つけることが可能であることを示した．

3. フェイクニュースの伝搬モデルの提案

本稿では，意見共有問題の枠組みの中で，EM アルゴリズムによる最尤推定を行うことで他人に対する信用度を動的に更新する手法を提案する．この手法により 1 章，2 章で述べた先行研究 [5] の問題点を解決する．そこで，本章

では，提案手法の土台となる意見共有問題の詳細と提案手法に関する説明を行う．

3.1 意見共有問題

エージェントの集合を $A = \{a_1, \ldots, a_N\}$ とし，各エージェントはエッジ E を持つものとする．このときの集団のネットワークを無向グラフ $G(A, E)$ で表す．各エージェント $i \in A$ は自分とエッジがあるエージェントの集合 $D_i = \{j : (i, j) \in E\}$ $(1 \leq |D_i| \leq N-1)$ を持つ．このとき，エージェント i は D_i 内の要素のエージェントとしか意見のやり取りができないと仮定する．以降，D_i の要素のエージェントをエージェント j と呼ぶことにする．

意見共有問題では，外部の客観的事実に相当する $\{b : b \in B = \{\mathrm{True}, \mathrm{False}\}\}$ に対する自分の意見を決めるための指標として，信念値 $P_i(b = \mathrm{True}) \in [0, 1]$ と呼ばれる値を持つ．これは，エージェント i が事実 b を本当 (True) であると思う確率を表す．逆に，エージェント i が事実 b を嘘 (False) であると思う確率は $1 - P_i(b = \mathrm{True})$ と表される．エージェント i はエージェント j の意見をもとに，ある信念値の初期値 P_i' から信念値を更新する．現在の信念値の更新回数を k とすると，エージェント i の信念値の更新式は以下の (1) 式の通りになる．

$$\begin{aligned} &P_i^k(b = \mathrm{T} | o_j = \mathrm{T}) \\ &= \frac{P(o_j = \mathrm{T} | b = \mathrm{T}) P_i^{k-1}(b = \mathrm{T})}{P(o_j = \mathrm{T} | b = \mathrm{T}) P_i^{k-1}(b = \mathrm{T}) + P(o_j = \mathrm{T} | b = \mathrm{F}) P_i^{k-1}(b = \mathrm{F})} \\ &= \frac{t_i P_i^{k-1}(z_1 = \mathrm{T})}{t_i P_i^{k-1}(b = \mathrm{T}) + (1 - t_i) P_i^{k-1}(b = \mathrm{F})} \end{aligned} \tag{1}$$

ここで，T は True，F は False を表す．そして，o_j はエージェント i がエージェント j から受け取った意見である．また，t_i とはエージェント i が全エージェント j の意見に対して持つ信用度を表し，$t_i \in [0.5, 1.0]$ と定義されている．(1) 式において，P_i^{k-1} が事前確率，t_i が尤度，P_i^k が事後確率に相当する．よって (1) 式はベイズの定理に基づいたベイズ更新となるが，この式と信用度 t_i には以下の問題点が存在する．

(i) 信用度が t_i であるため，j に依存しない．そのため，相手ごとに信用度が異なるように設定していない．

(ii) 信用度が $t_i \in [0.5, 1.0]$ であるため，エージェント i はすべてのエージェント j の意見を多少なりとも信用する．

(iii) (1) 式では，$P(o_j = \mathrm{T} | b = \mathrm{F}) = 1 - t_i$ と定義している．しかし，信用度を $P(o_j = \mathrm{T} | b = \mathrm{T}) = t_i$ と定義しているならば，$P(o_j = \mathrm{F} | b = \mathrm{T}) = 1 - t_i$ と定義されるべきである．したがって，(1) 式が誤っている．そのため，$P_i^k(b = \mathrm{F} | o_j = \mathrm{T})$，$P_i^k(b = \mathrm{T} | o_j = \mathrm{F})$，$P_i^k(b = \mathrm{F} | o_j = \mathrm{F})$ を正しく定義できない．

(iv) 信用度が常に一定の値で変化しない．

これらの問題を解決する手法を 3.2 節で説明する．

そして，エージェントの意見は，信念値がある閾値 σ 以上になる，あるいは閾値 $1 - \sigma$ 以下になることで決定される．ここで，k 回目の信念値の更新におけるエージェント i の意見 o_i^k は以下の (2) 式の通りに表される．

$$o_i^k = \begin{cases} \text{undetermined(意見が未定)} & \text{if} \quad 1 - \sigma < P_i^k < \sigma \\ \mathrm{True} & \text{if} \quad P_i^k \geq \sigma \\ \mathrm{False} & \text{if} \quad P_i^k \leq 1 - \sigma \\ o_i^{k-1} & \text{otherwise(意見が前の状態のまま)} \end{cases} \tag{2}$$

エージェント i は自分の意見を決めた瞬間に全てのエージェント j に自分の意見を伝え，エージェント j は受け取った意見から信念値を更新し，自分の意見を決定する．以上の流れを繰り返すことで，ベイズの定理に基づいた合理的な意見伝搬をモデル化している．

3.2 提案手法

本節では，3.1 節で述べた先行研究 [5] の 4 つの問題を解決するために EM アルゴリズムによる最尤推定を用いた信用度の更新方法を提案する．

3.2.1 意見行列の定義

本節では，2.3 節の Wang らの研究と提案手法の差異について述べる．Wang らは，M 人の報告者が L 個の事柄の真偽に関して行った報告結果を M 行 L 列の観測行列 SC により定義した．この観測行列では，ある報告者 $j (1 \leq j \leq M)$ がある事柄 $l (1 \leq l \leq L)$ を True と報告した場合は $S_j C_l = 1$，False と報告した場合は $S_j C_l = 0$ と定義している．本稿では，この観測行列をもとに，M 人のエージェント j が 1 つのニュースの真偽に関して行った報告結果を M 行 1 列の意見行列 SC により定義する．この意見行列では，エージェント j があるニュースについて True と伝えた場合は $S_j C_1 = 1$，False と伝えた場合は $S_j C_1 = 0$ とする．

3.2.2 信念値の定義の拡張

本稿では，信念値をエージェントがニュースの真実に相当する変数 z_1 を True であると思う確率であると定義し，k 回目の更新におけるエージェント i の信念値の更新式を (3) 式で表す．

$$\begin{aligned} &P_i^k(z_1 = \mathrm{T} | S_j C_1 = \mathrm{T}) \\ &= \frac{P(S_j C_1 = \mathrm{T} | z_1 = \mathrm{T}) P_i^{k-1}(z_1 = \mathrm{T})}{P(S_j C_1 = \mathrm{T} | z_1 = \mathrm{T}) P_i^{k-1}(z_1 = \mathrm{T}) + P(S_j C_1 = \mathrm{T} | z_1 = \mathrm{F}) P_i^{k-1}(z_1 = \mathrm{F})} \\ &= \frac{t_{ij} P_i^{k-1}(z_1 = \mathrm{T})}{t_{ij} P_i^{k-1}(z_1 = \mathrm{T}) + f_{ij} P_i^{k-1}(z_1 = \mathrm{F})} \end{aligned} \tag{3}$$

ここで，(1) 式と同様に，T は True，F は False を表す．また，$P_i^{k-1}(z_1 = \mathrm{F}) = 1 - P_i^{k-1}(z_1 = \mathrm{T})$ と定義する．そして，$P(S_j C_1 = \mathrm{T} | z_1 = \mathrm{T})$ はエージェント i がエージェント j に

対して持つ信用度 t_{ij} である．同様に，$P(S_jC_1 = \mathrm{T}|z_1 = \mathrm{F})$ はエージェント i がエージェント j に対して持つ不信度 f_{ij} である．これらの値域は，$t_{ij} \in [0.0, 1.0]$，$f_{ij} \in [0.0, 1.0]$ とする．これらは，エージェント i がエージェント j という人物ごとに信用し，エージェント j の意見を必ずしも信用しているわけではないという状況を表している．そして，(3) 式において，$P_i^{k-1}(z_1 = \mathrm{T})$ が事前確率，t_{ij} が尤度，$P_i^k(z_1 = \mathrm{T}|S_jC_1 = \mathrm{T})$ が事後確率に相当する．よって，(3) 式は (1) 式と同様，ベイズの定理に従ったベイズ更新であると言えるうえに，$P(S_jC_1 = \mathrm{T}|z_1 = \mathrm{F}) = f_{ij}$ と適切に定義しているため，(3) 式の定義は正しい．さらに，不信度 f_{ij} を定義したため，$P_i^k(z_1 = \mathrm{F}|S_jC_1 = \mathrm{T})$，$P_i^k(z_1 = \mathrm{T}|S_jC_1 = \mathrm{F})$，$P_i^k(z_1 = \mathrm{F}|S_jC_1 = \mathrm{F})$ を (3) 式と同様に正しく定義できる．以上から，3.1 節で説明した先行研究 [5] の問題のうち (i)，(ii)，(iii) を解決した．

3.3 EM アルゴリズムによる信用度の更新

はじめに，信用度の更新で使用する EM アルゴリズムの概要について説明する．EM アルゴリズムは最尤推定法の一つであり，期待値計算 (以下，E ステップと呼ぶ) と最大化計算 (以下，M ステップと呼ぶ) を交互に繰り返すことでパラメータの最尤推定値を求める．ここで，EM アルゴリズムの一般形を記す．観測変数を $\boldsymbol{X}$，推定対象のパラメータを $\boldsymbol{\theta}$，潜在変数を $\boldsymbol{Z}$，尤度関数を $L(\boldsymbol{X}, \boldsymbol{Z}; \boldsymbol{\theta})$ とすると，ある時点 t における E ステップは (4) 式，M ステップは (5) 式のように表すことができる．

$$Q(\boldsymbol{\theta}, \boldsymbol{\theta}^{(t)}) = E_{\boldsymbol{X},\boldsymbol{Z};\boldsymbol{\theta}^{(t)}}[\log L(\boldsymbol{X}, \boldsymbol{Z}; \boldsymbol{\theta})] \quad (4)$$

$$\boldsymbol{\theta}^{(t+1)} = \arg\max_{\boldsymbol{\theta}} Q(\boldsymbol{\theta}, \boldsymbol{\theta}^{(t)}) \quad (5)$$

(4) 式は一般に Q 関数と呼ばれる．

次に，提案手法における EM アルゴリズムを定式化する．はじめに，尤度関数 $L(\boldsymbol{X}, \boldsymbol{Z}; \boldsymbol{\theta})$ を (6) 式で定義する．このとき，観測変数 $\boldsymbol{X}$ は意見行列 S_jC_1，潜在変数 $\boldsymbol{Z}$ は真実 $z_1 \in \{\mathrm{True}, \mathrm{False}\}$，推定対象のパラメータ $\boldsymbol{\theta}$ は信用度 t_{ij} と不信度 f_{ij} である．

$$\begin{aligned} L(\boldsymbol{X}, \boldsymbol{Z}; \boldsymbol{\theta}) &= P(\boldsymbol{X}, \boldsymbol{Z}; \boldsymbol{\theta}) \\ &= \prod_{i=1}^{N} \left\{ \prod_{j=1}^{M} t_{ij}^{S_jC_1} (1 - t_{ij})^{(1-S_jC_1)} z_1 \right. \\ &\quad \left. + \prod_{j=1}^{M} f_{ij}^{S_jC_1} (1 - f_{ij})^{(1-S_jC_1)} (1 - z_1) \right\} \end{aligned} \quad (6)$$

また，対数尤度関数 $\log L(\boldsymbol{X}, \boldsymbol{Z}; \boldsymbol{\theta})$ は以下の (7) 式である．

$$\begin{aligned} &\log L(\boldsymbol{X}, \boldsymbol{Z}; \boldsymbol{\theta}) = \log P(\boldsymbol{X}, \boldsymbol{Z}; \boldsymbol{\theta}) \\ &= \sum_{i=1}^{N} \sum_{j=1}^{M} (S_jC_1 \log t_{ij} f_{ij} + (1 - S_jC_1) \log (1 - t_{ij}) (1 - f_{ij})) \end{aligned} \quad (7)$$

そして，Q 関数を (8) 式で定義する．

$$\begin{aligned} Q\left(\boldsymbol{\theta}, \boldsymbol{\theta}^{(t)}\right) &= P(\boldsymbol{Z}|\boldsymbol{X}; \boldsymbol{\theta}^{(t)}) \log P(\boldsymbol{X}, \boldsymbol{Z}; \boldsymbol{\theta}) \\ &= \sum_{i=1}^{N} \left\{ P_i^t \left(z_1 = 1|S_jC_1; \boldsymbol{\theta}^{(t)}\right) \right. \\ &\quad \times \left[\sum_{j=1}^{M} (S_jC_1 \log t_{ij} + (1 - S_jC_1) \log (1 - t_{ij})) \right] \\ &\quad + P_i^t \left(z_1 = 0|S_jC_1; \boldsymbol{\theta}^{(t)}\right) \\ &\quad \left. \times \left[\sum_{j=1}^{M} (S_jC_1 \log f_{ij} + (1 - S_jC_1) \log (1 - f_{ij})) \right] \right\} \end{aligned} \quad (8)$$

ここで，$P_i^t(z_1 = 1|S_jC_1; \boldsymbol{\theta}^{(t)})$ は $P_i^t(z_1 = \mathrm{T}|S_jC_1 = \mathrm{T})$ または $P_i^t(z_1 = \mathrm{T}|S_jC_1 = \mathrm{F})$ を表す．そして，$P_i^t(z_1 = 0|S_jC_1; \boldsymbol{\theta}^{(t)})$ は $P_i^t(z_1 = \mathrm{F}|S_jC_1 = \mathrm{T})$ または $P_i^t(z_1 = \mathrm{F}|S_jC_1 = \mathrm{F})$ を表す．

この E ステップにおいて，エージェント j の意見が未定の場合，意見行列の値が定義できないという問題がある．この問題を解決するために，意見が決定しているエージェント j の集合を A_i，意見が未定であるエージェント j の集合を $\overline{A_i}$ と定義し，エージェント j の意見に応じて A_i と $\overline{A_i}$ への割り当てを行う．例として，図 1 を示す．

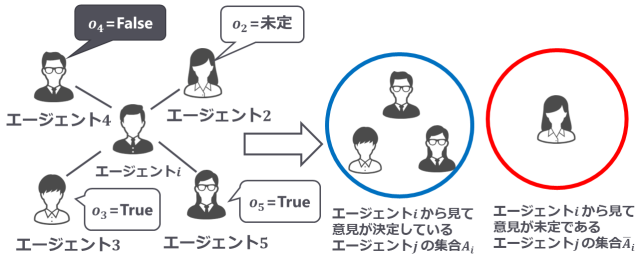

図 1 意見が決定したエージェント j の分割

この図において，エージェント j の要素に相当するエージェント 2 から 5 のうち，エージェント 2 の意見が未定である．そこで，エージェント 2 を $\overline{A_i}$ に，意見が決定している残りのエージェントを A_i に割り当てるという操作を行う．そして，意見が決定しているエージェントの集合 A_i に対して M ステップを適用する．具体的には，式 (8) を $\dfrac{\partial Q}{\partial t_{ij}} = 0, \dfrac{\partial Q}{\partial f_{ij}} = 0$ として偏微分し，図 1 で述べた処理を行うことで，式を変形すると，$t+1$ 回目の信用度の更新におけるエージェント i のエージェント j に対する信用度 $t_{ij}^{(t+1)}$，非信用度 $f_{ij}^{(t+1)}$ は以下の式で表すことができる．ここで K_i は t 回目の信用度の更新までに意見が決定しているエージェント j の総数である．

$$t_{ij}^{(t+1)} = \hat{t}_{ij} = \frac{\sum_{j\in A_i} P_i^t(z_1=1|S_jC_1;\boldsymbol{\theta}^t)}{\sum_{j=1}^{M} P_i^t(z_1=1|S_jC_1;\boldsymbol{\theta}^t)} \quad (9)$$

$$f_{ij}^{(t+1)} = \hat{f}_{ij} = \frac{K_i - \sum_{j\in A_i} P_i^t(z_1=1|S_jC_1;\boldsymbol{\theta}^t)}{M - \sum_{j=1}^{M} P_i^t(z_1=1|S_jC_1;\boldsymbol{\theta}^t)} \quad (10)$$

これらを (7) 式が収束するまで計算することにより，信用度 t_{ij} と非信用度 f_{ij} の最尤推定値を算出できる．以上から，3.1 節で説明した先行研究 [5] の問題のうち (iv) を解決したため，先行研究 [5] のすべての問題を解決した．

3.4 ネットワークにおける意見伝搬

本節では，提案手法を適用したときの意見伝搬のアルゴリズムについて説明する．はじめに，あるエージェント i に着目した意見伝搬を考える．このとき，エージェント i は，はじめに (I) を行う．次に，(II) を行ったあとに (III) を行う，または (III) を行ったあとに (II) という手順にしたがって意見を伝える．

(I) エージェント i の信念値が閾値 σ 以上になる，あるいは閾値 $1-\sigma$ 以下になることで意見を決定する．その意見を M 人のエージェント j に伝える．

(II) エージェント i が M 人のエージェント j の意見と信念値と更新前の信用度 t_{ij} と不信度 f_{ij} を使うことで，EM アルゴリズムにより信用度と不信度を更新する．

(III) 信用度と不信度をもとにエージェント i が信念値を更新する．

今回は，(II) を行ったあとに (III) を行うように設定した．つまり，エージェント i は信用度と不信度を更新したあと，更新した信用度と不信度を使って信念値を更新する．この設定は，個人が他人の意見を熟慮したあとにその人物の信用度を決定することで，意見を決定するという状況を反映している．この手順 (III)〜(III) をネットワーク上の全てのエージェントが行う．ここで，以上の手順を行うタイミングに関して以下の 3 つの状況が想定される．

(A) 全エージェントが同時に (I)〜(III) を行う．

(B) エージェントごとに異なるタイミングで (I)〜(III) を行う．

(C) 影響力を持つ人物や友人が多い人物など，意見を発信する優先度が高いと考えられるエージェントがはじめに (I)〜(III) を行う．

今回は (A) を採用した．

以上の流れを表したものとして，提案手法のアルゴリズムを以下のアルゴリズム 1 として示す．ここで，i，j はエージェント i とエージェント j の番号，K は信念値の最大更新回数，T は信用度の最大更新回数，σ は意見を決定するための閾値，ϵ は (6) 式の尤度関数の収束を判定するための閾値，δ は意見が決定したエージェントの割合 (以下，意見決定率と呼ぶ．)，o_i はエージェント i の意見，L^t は t 回の信用度の更新のときの対数尤度関数を表す．

アルゴリズム 1
EM アルゴリズムにより個人間の信用度と不信度を動的に更新したときの意見伝搬

```
ネットワークトポロジーを設定.
真実 z_1 の真偽を設定.
閾値 σ, ϵ, δ を設定.
for i = 1 to N do
    信念値の初期値 P_i^0 を正規乱数で設定
    信用度の初期値 t_ij^0，非信用度 f_ij^0 の初期値を一様乱数で設定
end for
while 意見決定率 ≤ δ or k ≤ K do
    for i = 1 to N do
        if P_i^k ≥ σ then
            o_i = True
            意見 o_i=True をエージェント j に発信
        else if P_i^k ≤ 1 − σ then
            o_i = False
            意見 o_i=False をエージェント j に発信
        end if
    end for
    for i = 1 to N do
        for t = 0 to T do
            if t = 0 then
                対数尤度関数 L^0 = 0 として初期化
            end if
            while L^t − L^{t−1} > ϵ do
                for j = 1 to M do
                    エージェント j の意見を意見行列 SC に格納
                    対数尤度関数 L^t を (7) 式により計算
                    信用度 t_ij^t を (9) 式により更新
                    非信用度 f_ij^t を (10) 式により更新
                end for
            end while
        end for
        信念値 P_i^k を更新
    end for
end while
```

4. 実験

本章で述べる設定にしたがって評価実験を行った．

4.1 ネットワークトポロジー

以下の 3 つのネットワークトポロジーをソーシャルネットワークとみなして，フェイクニュースの伝搬をモデル化する．

- スケールフリーネットワーク [12]: 多くのノードとつながりを持つハブと呼ばれるノードが一部存在するネットワーク．
- スモールワールドネットワーク [13]: ネットワーク内

の任意の 2 つのノード間の距離が小さくなるネットワーク

- ランダムネットワーク [14]: ネットワーク内の任意の 2 つのノードが確率 p で結ばれるネットワーク

4.2 比較対象

3.1 節で説明した信用度 t_i，3.2 節で説明した信用度 t_{ij} と不信度 f_{ij} に関して以下の通りに設定した．

- 信用度 t_i を [0.5, 1.0] の範囲の一様分布で設定し，その値を変化させない．(以下，先行手法と呼ぶ)
- 信用度 $t_{ij} \in [0.0, 1.0]$ と不信度 $f_{ij} \in [0.0, 1.0]$ を一様乱数で初期化し，EM アルゴリズムにより最尤推定することでこれらを更新する．(以下，提案手法と呼ぶ)

4.3 評価指標

先行手法と提案手法のそれぞれに対して, 以下の割合を求めた．

- 真実 z_1 と同じ意見を持っているエージェントの割合 (以下，正解率と呼ぶ)
- 真実 z_1 と異なる意見を持っているエージェントの割合 (以下，不正解率と呼ぶ)
- 意見が未定であるエージェントの割合 (以下，未定率と呼ぶ)

ここで，アルゴリズム 1 における意見決定率 δ とは，正解率と不正解率の合計値を表す．

これらの指標から先行手法と提案手法における意見伝搬の性質に関して考察する．

4.4 パラメータの設定

今回の実験では，エージェント数 $N = 1000$，信念値の最大更新回数 $K = 500$，信用度と不信度の最大更新回数 $T = 15$，意見を決定するための閾値 $\sigma = 0.8$，(6) 式の尤度関数の収束を判定する閾値 $\epsilon = 0.00001$，意見決定率 $\delta = 0.8$ と設定した．また，信念値の初期値 P_i^0 を真実 $z_1 = \text{True}$ のときは $\mathcal{N}(0.55, 0.10)$ の正規乱数，$z_1 = \text{False}$ のときは $\mathcal{N}(0.45, 0.10)$ の正規乱数とした．信用度の初期値 t_{ij}^0 と不信度の初期値 f_{ij}^0 はそれぞれ一様乱数で設定した．そして，これらのパラメータを 4.1 節で説明した 3 つのネットワークトポロジーに対して適用し，先行手法と提案手法において正解率，不正解率，未定率を算出した．

5. 結果と考察

5.1 結果

4.4 節の設定にしたがってシミュレーションを行った．その結果を以下の表 1 に示す．ここで，シミュレーションはニュースの真実 $z_1 = \text{True}$ のときで 50 回，$z_1 = \text{False}$ のときで 50 回の合計 100 回行った．そして，正解率，不正解率，未定率の値は 100 回のシミュレーションの平均値である．また，表中の ScaleFree はスケールフリーネットワーク，SmallWorld はスモールワールドネットワーク，Random はランダムネットワークを表す．

手法	トポロジー	正解率	不正解率	未定率
提案手法	ScaleFree	70.6%	18.7%	10.7%
	SmallWorld	64.9%	19.7%	15.4%
	Random	73.9%	19.5%	6.7%
先行手法	ScaleFree	65.5%	20.8%	13.7%
	SmallWorld	62.0%	21.7%	16.3%
	Random	70.7%	21.5%	7.7%

表 1 ネットワークトポロジーに応じた先行手法と提案手法の正解率，不正解率，未定率の比較

表 1 の結果を見ると，いずれの手法でもネットワークトポロジーに関わらず正答率が向上し，提案手法ではその傾向が顕著であることがわかる．以降の節でこの理由について考察する．

5.2 考察

今回, 信念値の初期値は真実 $z_1 = \text{True}$ のときは $\mathcal{N}(0.55, 0.10)$ の正規乱数，$z_1 = \text{False}$ のときは $\mathcal{N}(0.45, 0.10)$ の正規乱数とした．そのため，いずれの手法でも真実と同じ意見を共有する可能性が高くなり，初期段階では意見が未定であるエージェントが多くなる．このとき，提案手法では先行手法と比較して，ネットワークトポロジーに関わらず未定率が低いため，正答率が向上したと考えられる．5.2.1 節で，提案手法において正答率が向上した理由を詳細に考察する．

5.2.1 信念値の分布と信用度，不信度の分布からの考察

はじめに，先行手法と提案手法を適用した各ネットワークトポロジーにおいて，シミュレーション終了時の信念値の分布を以下の図 2，図 3，図 4 に示す．ここで，横軸は信念値，縦軸はエージェントの数を表し，その度数は各真実に対してシミュレーション 50 回行ったときの合計である．これらの図は，エージェントの信念値が 0 に近いほど意見が False 側の意見を持ち，1 に近いほど True 側の意見を持つという状況を表している．図 2，図 3，図 4 から，先行手法ではネットワークトポロジーに関わらず真実 $z_1 = \text{True}$ のときは，正しい意見を持つエージェントが多く，$z_1 = \text{False}$ のときは，意見の二極化が起こった．提案手法では，ネットワークトポロジーに関わらずニュースの真実 $z_1 = \text{True}$ のときは正しい意見を持っているエージェントの数が最も多く，$z_1 = \text{False}$ のときは，先行手法と比較すると意見に多様性が生じる結果となった．

はじめに，提案手法と比較すると先行手法では正しい意見を共有できなかった理由を考察する．まず，先行手法では信用度が値域が [0.5, 1.0] の範囲で定義され，一様分布にしたがって設定し，その分布は変化しない．そのため，相手の意見を信用して信念値を更新するため，誤った意見も受け入れる．その結果，真実が True の場合は正しい意見を共有できるが，真実が False の場合は True の意見を受け入れるため，意見の二極化が起こり，正しい意見が共有されにくくなったと考えられる．

次に，提案手法において正しく意見を共有できた理由を考察する．そのために，信用度と不信度の分布を図 5 に示す．ここで，横軸は信用度あるいは不信度，縦軸はエージェントの数を表し，その度数はシミュレーションを 100 回行ったときの合計である．図 5 から，不信度は，ネットワークトポロジーに関わらず値が 0 となるエージェントが最も多くなり，信用度は高い値を持つエージェントが多くなることが分かる．これらの分布をもとに，提案手法において正しい意見が共有された理由を考察する．真実 z_1 が True の場合，信念値 $P_i^k(z_1 = \mathrm{T}|S_jC_1 = \mathrm{T})$ と $P_i^k(z_1 = \mathrm{T}|S_jC_1 = \mathrm{F})$ の定義から，信用度 t_{ij} が意見を決定するための重要な変数となる．具体的には，他人の意見が True のとき，エージェントは (3) 式を適用する．このとき，信用度が高い値を持つエージェントが多いことから，他人の意見を信用するため，エージェントは True という意見を持ちやすくなる．そして，他人の意見が False のとき，エージェントは

$$P_i^k(z_1 = \mathrm{T}|S_jC_1 = \mathrm{F}) = \frac{(1-t_{ij})P_i^{k-1}(z_1 = \mathrm{T})}{(1-t_{ij})P_i^{k-1}(z_1 = \mathrm{T}) + (1-f_{ij})P_i^{k-1}(z_1 = \mathrm{F})}$$

を適用する．このとき，信用度が高い値を持つエージェントが多いことから，真実と異なる意見を持つ他人を信用しないという状況が起こるため，エージェントは True という意見を持ちやすくなる．よって，図 2，図 3，図 4 の (a) の信念値の分布も考慮すると，真実 $z_1 = \mathrm{True}$ のとき，つまり，あるニュースの真実が本当であった場合，本当であると伝えた他人は信用し，嘘だと伝えた他人は信用しないことで，正しい意見が拡散されやすくなることが明らかとなった．

次に，真実 z_1 が False の場合，$P_i^k(z_1 = \mathrm{F}|S_jC_1 = \mathrm{T})$ と $P_i^k(z_1 = \mathrm{F}|S_jC_1 = \mathrm{F})$ の定義から不信度 f_{ij} が意見を決定するための重要な変数となる．具体的には，他人の意見が True のとき，エージェントは

$$P_i^k(z_1 = \mathrm{F}|S_jC_1 = \mathrm{T}) = \frac{f_{ij}P_i^{k-1}(z_1 = \mathrm{T})}{t_{ij}P_i^{k-1}(z_1 = \mathrm{T}) + f_{ij}P_i^{k-1}(z_1 = \mathrm{F})}$$

を適用する．今回，不信度 f_{ij} が 0 であるエージェントが最も多いことから，エージェントは False の意見を持ちやすくなる．また，他人の意見が False のとき，

$$P_i^k(z_1 = \mathrm{F}|S_jC_1 = \mathrm{F}) = \frac{(1-f_{ij})P_i^{k-1}(z_1 = \mathrm{F})}{(1-t_{ij})P_i^{k-1}(z_1 = \mathrm{T}) + (1-f_{ij})P_i^{k-1}(z_1 = \mathrm{F})}$$

を適用する．このとき，不信度 f_{ij} が 0 であるエージェントが最も多いことから，真実と同じ意見を持つ他人を信用するため，エージェントは False の意見を持ちやすくなる．したがって，図 2，図 3，図 4 の (c) の信念値の分布を考慮すると，真実 $z_1 = \mathrm{False}$ のとき，つまり，あるニュースがフェイクニュースであった場合，フェイクニュースであると伝えた人物の意見は信用し，そうでない人物の意見を信用しないことで，正しい意見が拡散されやすくなることが明らかとなった．

以上より，提案手法により，個人間の信用度を動的に更新することで，信頼できる人物と信頼できない人物を区別できるようになったため，正しい意見が拡散されやすくなることを示した．

6. おわりに

本稿では，EM アルゴリズムを適用することで個人間で信用度を動的に更新する手法に基づいて，フェイクニュースにおける意見の伝搬過程を考慮したマルチエージェントモデルを提案した．

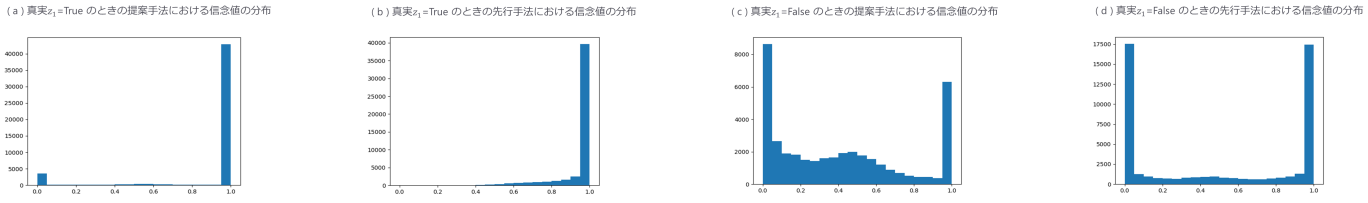

図 2 スケールフリーネットワークにおける先行手法と提案手法の信念値の分布

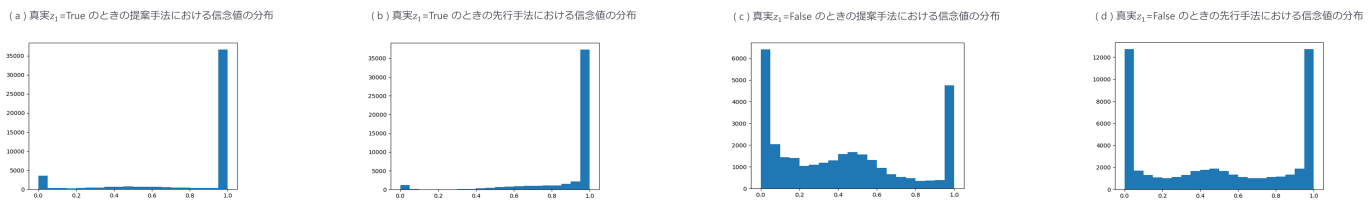

図 3 スモールワールドネットワークにおける先行手法と提案手法の信念値の分布

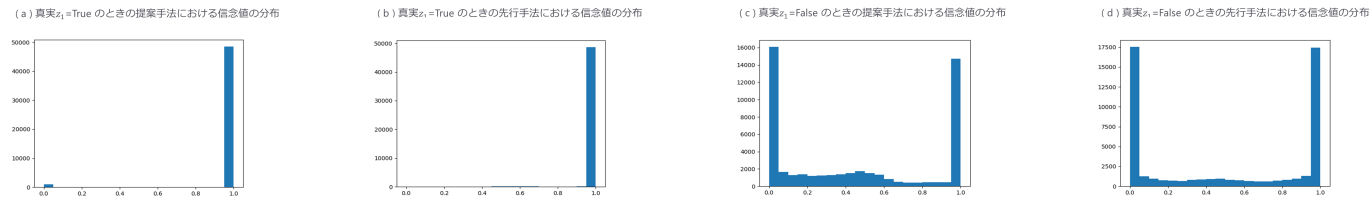

図 4 ランダムネットワークにおける先行手法と提案手法の信念値の分布

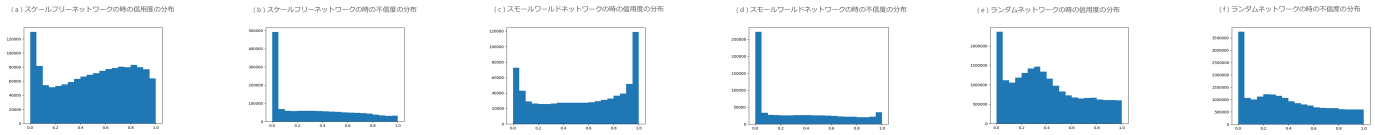

図 5 信用度と不信度の分布

先行研究 [5] では，誤った意見の存在下でも，正しい意見を共有できる手法を提案した．しかし，この手法には複数の問題が存在する．はじめに，他人の意見への信用度が一定であり，かつ他人に応じて信用度が区別されていないという問題がある．また，この手法における信用度の定義では，他人の意見を多少なりとも信用することが前提となっているため，フェイクニュースの伝搬を表現できる手法ではない．そして，ベイズの定理に基づいた意見の更新式が誤っているという問題も存在する．そこで，本稿ではこれらの問題点を解決し，評価実験を行った．その結果，ネットワークの形態に関わらず意見が決まりやすくなり，かつ正しい意見が拡散されやすくなることを示した．今後の課題として以下のものがある．

- 提案手法では，ある一人の他人の意見を受け取り，その意見を伝えた他人に対する信用度の更新を行っている．しかし，ある人物から意見を聞いたあとに，より信用できる別の人物からの異なる意見を聞いた結果，はじめに意見を聞いた人物の意見を信用しない場合があると考えられる．そのため，複数人の意見を考慮した信用度の更新を行う手法を提案する．
- 図 5 を見ると，不信度の値が 1 になっているエージェントの数が少ない．不信度の定義から考えると，真実が False のとき他人の意見が True ならば他人を信用しないという人物が少ないことになる．この原因を分析する．

参考文献

[1] Nyhan, B. and Reifler, J.: When corrections fail: The persistence of political misperceptions, *Political Behavior*, Vol. 32, No. 2, pp. 303–330 (2010).
[2] Quattrociocchi, W., Scala, A. and Sunstein, C. R.: Echo chambers on Facebook, *Available at SSRN 2795110* (2016).
[3] Rubin, V., Conroy, N., Chen, Y. and Cornwell, S.: Fake news or truth? using satirical cues to detect potentially misleading news, *Proceedings of the second workshop on computational approaches to deception detection*, pp. 7–17 (2016).
[4] Yang, S., Shu, K., Wang, S., Gu, R., Wu, F. and Liu, H.: Unsupervised fake news detection on social media: A generative approach, *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 33, pp. 5644–5651 (2019).
[5] Pryymak, O., Rogers, A. and Jennings, N. R.: Efficient opinion sharing in large decentralised teams, *Proceedings of the 11th International Conference on Autonomous Agents and Multiagent Systems-Volume 1*, International Foundation for Autonomous Agents and Multiagent Systems, pp. 543–550 (2012).
[6] Yao, Y., Viswanath, B., Cryan, J., Zheng, H. and Zhao, B. Y.: Automated crowdturfing attacks and defenses in online review systems, *Proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security*, ACM, pp. 1143–1158 (2017).
[7] Vosoughi, S., Roy, D. and Aral, S.: The spread of true and false news online, *Science*, Vol. 359, No. 6380, pp. 1146–1151 (2018).
[8] Ferrara, E., Varol, O., Davis, C., Menczer, F. and Flammini, A.: The rise of social bots, *Communications of the ACM*, Vol. 59, No. 7, pp. 96–104 (2016).
[9] 白井嵩士，　榊剛史，鳥海不二夫，篠田孝祐，風間一洋，野田五十樹，沼尾正行，　栗原聡：Twitter におけるデマツイートの拡散モデルの構築とデマ拡散防止モデルの推定，人工知能学会全国大会予稿集, IC3-OS-12-1 (2012).
[10] Glinton, R. T., Scerri, P. and Sycara, K.: Towards the understanding of information dynamics in large scale networked systems, *Information Fusion, 2009. FUSION'09. 12th International Conference on*, IEEE, pp. 794–801 (2009).
[11] Wang, D., Kaplan, L., Le, H. and Abdelzaher, T.: On truth discovery in social sensing: A maximum likelihood estimation approach, *Proceedings of the 11th international conference on Information Processing in Sensor Networks*, ACM, pp. 233–244 (2012).
[12] Barabási, A.-L. and Albert, R.: Emergence of scaling in random networks, *science*, Vol. 286, No. 5439, pp. 509–512 (1999).
[13] Watts, D. J. and Strogatz, S. H.: Collective dynamics of ' small-world ' networks, *nature*, Vol. 393, No. 6684, p. 440 (1998).
[14] Erdős, P. and Rényi, A.: On the evolution of random graphs, *Publ. Math. Inst. Hung. Acad. Sci*, Vol. 5, No. 1, pp. 17–60 (1960).