

# ソーシャルメディアでの信頼ネットワーク進化における 媒介者アクティビティ効果の分析

八田 圭斗<sup>1</sup> 熊野 雅仁<sup>2</sup> 木村 昌弘<sup>2</sup> 齊藤 和巳<sup>3</sup> 大原 剛三<sup>4</sup> 元田 浩<sup>5</sup>

**概要:** 本論文では、ソーシャルメディアにおける信頼ネットワークの進化過程を媒介者の観点から分析する。そのために、媒介者が存在するユーザ間の信頼リンク生成ダイナミクスを分析するモデルとして、2つの確率過程モデル A-ME と A-MAE を提案する。ここに、A-ME モデルは信頼ネットワーク進化における媒介者タイプ情報の効果を分析し、A-ME モデルの拡張である A-MAE モデルは、さらに媒介者アクティビティ共起情報の効果を分析する。Epinions の実データを分析し、媒介者が存在するユーザ間における近い将来の信頼リンク予測において、A-ME モデルはベースライン法よりも精度が高く、さらに A-MAE モデルは A-ME モデルよりも大幅に精度が高いことを示し、媒介者タイプ情報と媒介者アクティビティ共起情報を組み込むことは、信頼リンク生成ダイナミクスのモデル化において有効であることを実証する。

## 1. はじめに

ソーシャルメディアが現れたことで、人々間の信頼関係についての大規模ネットワークが収集可能となってきた。ここに、ユーザ  $u$  からユーザ  $v$  に信頼リンクが生成されたということは、 $v$  は  $u$  にとって信頼できるユーザであり、 $u$  は  $v$  が与える様々な情報やオピニオンを受け入れたり共有したりする傾向があることを意味している。信頼ネットワーク構造は、社会的見地から重要であるだけでなく、オンライン世界において信頼できる情報を効率良く同定するのにも利用可能である。また、信頼ネットワークはソーシャルネットワークの一種であり、情報拡散やオピニオン形成において重要な役割を果たしている [2], [6], [7], [8], [14]。したがって、近年、信頼リンクやソーシャルリンクを予測する問題の研究に注目が集まってきている [4], [5], [10], [11], [12], [13], [15], [16]。

ユーザ  $u$  が他のユーザ  $v$  に信頼リンクを生成する理由は様々であるはずだが、信頼リンク生成においてこれら 2 ユーザ  $u$  と  $v$  の間を媒介するユーザの役割を調べることは、最も基本的な研究課題と考えられる。実際、Leskovec ら [9] は、符号つきソーシャルネットワークにおけるリンク  $(u, v)$  の符号 (ポジティブかネガティブ) を推定するための基本特徴量として、リンク  $(u, v)$  を含むトライアド構造

を用いた。本論文では、媒介者の観点から信頼ネットワークの進化過程を分析する。ここに、ユーザ  $u$  からユーザ  $v$  への媒介者  $w$  とは、「 $w$  は  $u$  からかまたは  $u$  への信頼リンクを持ち、さらに、 $w$  は  $v$  からかまたは  $v$  への信頼リンクを持つ。」というユーザとして定義される。さて、ユーザ  $u$  からユーザ  $v$  への媒介者には 4 つのタイプが存在する。例えば、タイプ 1 媒介者  $w$  とは、 $u$  から  $w$  への信頼リンクと  $w$  から  $v$  への信頼リンクを持つものである (図 1 参照)。異なるタイプの媒介者は信頼リンク生成に対する影響も異なり得ると、自然に推測される。したがって我々は、媒介者タイプの観点から信頼リンク生成における媒介者の影響を分析する。

ところで、いくつかのソーシャルメディアサイトは、信頼リンクを結ぶ環境とともにアクティビティを実行できるようなサービスも提供している。そこでは、ユーザは与えられた一連のアクティビティ群から一つのものを選択し実行することが可能である。例として製品レビューサイトを考えると、ユーザ達は、与えられた製品群内の製品に対して、そのレビューを投稿したりその評点を与えたりすること (ユーザアクティビティ) が可能であるし、また、互いに信頼リンクを生成すること (社会交流) も可能である。Crandall ら [3] は、オンラインコミュニティにおいて社会交流と行動類似性との間に相互関係が存在することを実証した。この研究より、ソーシャルメディアサイトにおいて、「1) 誰への信頼リンクが生成されるか」と「2) どのアクティビティが実行されるか」との間には、密接な関係があることが示唆される。

<sup>1</sup> 龍谷大学大学院理工学研究科電子情報学専攻

<sup>2</sup> 龍谷大学理工学部電子情報学科

<sup>3</sup> 静岡県立大学経営情報学部経営情報学科

<sup>4</sup> 青山学院大学理工学部情報テクノロジー学科

<sup>5</sup> 大阪大学産業科学研究所

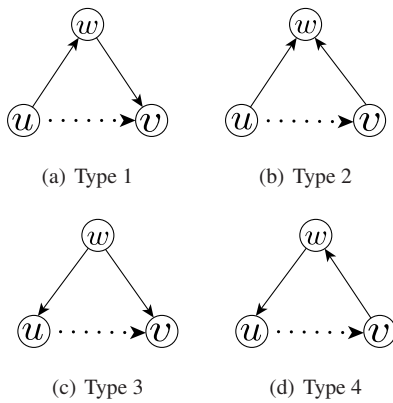


図1 ノード  $u$  からノード  $v$  への媒介者  $w$  の4タイプ  
Fig. 1 Four types of mediator  $w$  from node  $u$  to node  $v$ .

本論文では、ソーシャルメディアにおける信頼ネットワークの進化過程を説明できるモデルの構築を目指して、特に、媒介者の観点からアクティビティ情報の効果を分析することを目的とする。そのために、媒介者が存在するユーザ間の信頼リンク生成ダイナミクスを分析するモデルとして、2つの確率過程モデル A-ME と A-MAE を提案する。ここに、A-ME モデルは信頼ネットワーク進化における媒介者タイプ情報の効果を分析するものであり、A-ME モデルの拡張である A-MAE モデルは、さらに媒介者アクティビティ共起情報の効果を分析するものである。A-ME モデルと A-MAE モデルを用いて、製品レビューと消費者レポートに関するソーシャルメディアサイトである Epinions<sup>\*1</sup> の実データを分析する。まず、媒介者が存在する2ユーザ間における近い将来の信頼リンク予測に関して、A-ME モデルと A-MAE モデルを評価する。そして、A-ME モデルはベースライン法よりも精度が高いことを確認し、媒介者タイプ情報を組み込むことは、信頼リンク生成ダイナミクスのモデル化において効果があることを実証する。また A-MAE モデルは A-ME モデルよりも大幅に精度が高いことを示し、媒介者アクティビティ共起情報とそれに基づく媒介者タイプ情報を組み込むことは、信頼リンク生成ダイナミクスのモデル化において非常に有効であることを実証する。

## 2. 準備

本研究では、ソーシャルメディアサイトにおいてユーザが行う、信頼リンク生成とアクティビティについて考える。まず、本論文を通して使用する記号法を導入する。

正の整数  $t$  に対して、期間  $I_t = (t_0 + (t-1)\Delta t, t_0 + t\Delta t]$  内に生成された信頼ネットワークを  $G_t = (V, E_t)$  とする。ここに、 $V$  はノード集合であり、初期時刻  $t_0$  におけるソーシャルメディアサイト内のユーザ全体の集合を表している。 $E_t \subset V \times V$  は、期間  $I_t$  内に生成された信頼リンク全体の集

合である。 $\Delta t$  はあらかじめ指定する正定数であり、期間の長さを表している。ここでは、自己リンクや多重リンクは存在しないとする。 $\bar{G}_t = (V, \bigcup_{s=1}^t E_s)$  は時刻  $t_0 + t\Delta t$  でのユーザ集合  $V$  における信頼ネットワークを表すこと、また、 $s \neq s'$  ならば  $E_s \cap E_{s'} = \emptyset$  であることに注意する。さて、次の期間  $I_{t+1}$  内に生成される信頼リンク集合  $E_{t+1}$  を予測することを考えよう。 $V \times V$  の部分集合  $C_{t+1}$  を

$$C_{t+1} = (V \times V) \setminus \{(v, v) | v \in V\} \setminus \bigcup_{s=1}^t E_s$$

と定義する。このとき、 $E_{t+1} \subset C_{t+1}$  であり、 $\bigcup_{s=1}^t E_s \cap C_{t+1} = \emptyset$  であることは容易に示される。従って、 $C_{t+1}$  を期間  $I_{t+1}$  における信頼リンク候補集合と呼ぶ。我々は、任意の信頼リンク候補  $(u, v) \in C_{t+1}$  に対して、ノード  $u$  からノード  $v$  への信頼リンクが期間  $I_{t+1}$  内に生成されるかどうかを調べる。

ソーシャルメディアサイトでは、ユーザは  $K$  種類のアクティビティを選択し実行できるものとする。任意のノード  $u \in V$  と正の整数  $t$  に対して、期間  $I_t$  におけるノード  $u$  のアクティビティベクトルを

$$A_t(u) = (A_{t,1}(u), \dots, A_{t,K}(u))$$

とする。ただし、各  $k$  について  $A_{t,k}(u)$  は、ユーザ  $u$  が期間  $I_t$  内にアクティビティ  $k$  を選択し実行すれば  $A_{t,k}(u) = 1$  であり、そうでないならば  $A_{t,k}(u) = 0$  である。

本論文では、信頼リンク生成ダイナミクスにおける媒介者の役割を調べることを目的とする。したがって、信頼リンク候補集合  $C_{t+1}$  全体ではなく、期間  $I_t$  に媒介者  $w \in V$  を有する信頼リンク候補  $(u, v) \in C_{t+1}$  からなる集合  $C_{t+1}^* \subset C_{t+1}$  に焦点を当てる。集合  $C_{t+1}^*$  を期間  $I_{t+1}$  における許容信頼リンク候補集合と呼び、その要素  $(u, v) \in C_{t+1}^*$  を期間  $I_{t+1}$  における許容信頼リンク候補と呼ぶ。我々は、任意の許容信頼リンク候補  $(u, v) \in C_{t+1}^*$  に対して、期間  $I_{t+1}$  内にノード  $u$  からノード  $v$  へ信頼リンクが生成される、すなわち  $(u, v) \in E_{t+1}$  である確率  $P_{t+1}(u, v)$  をモデル化することを考える。ここに、期間  $I_t$  において、ノード  $u$  とノード  $w$  の間に信頼リンクが生成され、かつ、ノード  $v$  とノード  $w$  の間に信頼リンクが生成されたとき、ノード  $w$  は期間  $I_t$  におけるノード  $u$  からノード  $v$  への媒介者と呼ぶ。そして、期間  $I_t$  におけるノード  $u$  からノード  $v$  への媒介者全体の集合を  $M_t(u, v)$  とする。 $M_t(u, v) = M_t(v, u)$  に注意する。ところで、期間  $I_t$  におけるノード  $u$  からノード  $v$  への媒介者  $w$  は、次のように4つのタイプに分類される。 $(u, w), (w, v) \in E_t$  ならば  $w$  をタイプ1媒介者と呼び(図1(a)参照)、それら全体の集合を  $M_{t,1}(u, v)$  とし、 $(u, w), (v, w) \in E_t$  ならば  $w$  をタイプ2媒介者と呼び(図1(b)参照)、それら全体の集合を  $M_{t,2}(u, v)$  とし、 $(w, u), (w, v) \in E_t$  ならば  $w$  をタイプ3媒介者と呼び(図1(c)参照)、それら全体の集合を  $M_{t,3}(u, v)$  とし、 $(w, u), (v, w) \in E_t$  ならば  $w$  をタイプ4媒介者と呼び

\*1 <http://www.epinions.com>

(図 1(d) 参照), それら全体の集合を  $\mathcal{M}_{t,4}(u, v)$  とする.

$$\mathcal{M}_t(u, v) = \bigcup_{i=1}^4 \mathcal{M}_{t,i}(u, v)$$

であること, また,  $\mathcal{M}_{t,1}(u, v) \neq \mathcal{M}_{t,1}(v, u)$ ,  $\mathcal{M}_{t,4}(u, v) \neq \mathcal{M}_{t,4}(v, u)$  であることに注意する. 我々は, さらに, アクティビティの観点から信頼リンク生成ダイナミクスにおける媒介者の役割を分析する. したがって, 任意の許容信頼リンク候補  $(u, v) \in C_{t+1}^*$  に対して信頼リンク生成確率  $P_{t+1}(u, v)$  をモデル化するとき, アクティビティ情報を組み込むことを考える.

### 3. A-ME モデルと A-MAE モデル

任意の許容信頼リンク候補  $(u, v) \in C_{t+1}^*$  について, 期間  $I_{t+1}$  内に信頼リンク  $(u, v)$  が生成される, すなわち,  $(u, v) \in E_{t+1}$  である確率  $P_{t+1}(u, v)$  のモデル化を考える. 定義より,  $I_t$  におけるノード  $u$  からノード  $v$  への媒介者は必ず存在することに注意する. 信頼リンク生成ダイナミクスに対して, 媒介者のタイプ情報およびアクティビティ情報の効果を分析するために,  $P_{t+1}(u, v)$  の 2 つの自然なモデル A-ME と A-MAE を提案する. A-ME モデルは媒介者のタイプ情報の効果を分析するものであり, A-MAE モデルは, A-ME モデルの拡張であって, さらに媒介者アクティビティ共起情報の効果を分析するモデルである. 提案モデルでは, 信頼リンク形成における各媒介者  $w$  の影響の度合いを  $w$  の媒介者価値と呼び, 正数  $g(w)$  でスコアリングする<sup>\*2</sup>.

#### 3.1 A-ME モデル

媒介者の存在は信頼リンクの生成に影響を及ぼすと考えられる. さらに, 媒介者の影響の強さはそのタイプに依存して異なると推測できる. したがって, 媒介者タイプの観点から, 信頼リンク生成における媒介者の影響を分析するために, 任意の  $(u, v) \in C_{t+1}^*$  に対し確率  $P_{t+1}(u, v)$  をロジスティック回帰モデルにより

$$P_{t+1}(u, v) = \frac{1}{1 + \exp(-\boldsymbol{\phi} \cdot \mathbf{y}_t(u, v))} \quad (1)$$

とモデル化することを提案する. ここで,  $\boldsymbol{\phi} \in \mathbf{R}^5$  はパラメータベクトル

$$\boldsymbol{\phi} = (\phi_0, \phi_1, \phi_2, \phi_3, \phi_4),$$

$\mathbf{y}_t(u, v)$  は時刻  $t_0 + t\Delta t$  における  $(u, v)$  の特徴ベクトル

$$\mathbf{y}_t(u, v) = (1, y_{t,1}(u, v), y_{t,2}(u, v), y_{t,3}(u, v), y_{t,4}(u, v)),$$

また,  $\boldsymbol{\phi} \cdot \mathbf{y}_t(u, v)$  はベクトル  $\boldsymbol{\phi}$  とベクトル  $\mathbf{y}_t(u, v)$  の内積

<sup>\*2</sup> 媒介者  $w$  の媒介者価値のスコアリングに関しては,  $w$  が得た信頼リンク数,  $w$  が作成した信頼リンク数,  $w$  が実行したアクティビティ数, また,  $w$  が行ったレビューに対する評点など, 様々なものが考えられる. 実験では, よく用いられる 2 つのスコアリング法を採用した(4.2.1 節参照).

$$\boldsymbol{\phi} \cdot \mathbf{y}_t(u, v) = \phi_0 + \sum_{i=1}^4 \phi_i y_{t,i}(u, v)$$

である. ただし,

$$y_{t,i}(u, v) = \sum_{w \in \mathcal{M}_{t,i}(u, v)} g(w) \quad (i = 1, 2, 3, 4)$$

である.  $y_{t,i}(u, v)$  は, ノード  $u$  からノード  $v$  へのタイプ  $i$  媒介者達全体の影響度を表していることに注意する. 我々は, 媒介者タイプの観点から信頼リンク生成ダイナミクスを分析する, このような確率モデルを A-ME モデルと呼ぶ.

#### 3.2 A-MAE モデル

信頼リンク生成における媒介者の影響の度合いは, アクティビティにも依存すると考えられる. 許容信頼リンク候補  $(u, v) \in C_{t+1}^*$  に対して, 期間  $I_t$  におけるノード  $u$  からノード  $v$  への次のような 2 人の媒介者  $w_k$  と  $w_\ell$  を考えよう. ここに, ノード  $u$ , ノード  $v$  およびノード  $w_k$  は期間  $I_t$  においてアクティビティ  $k$  を実行(すなわち, アクティビティ  $k$  で共起)し, ノード  $u$ , ノード  $v$  およびノード  $w_\ell$  は期間  $I_t$  においてアクティビティ  $\ell$  を実行(アクティビティ  $\ell$  で共起)したとする. すなわち,  $A_{t,k}(u) = A_{t,k}(v) = A_{t,k}(w_k) = 1$  であり,  $A_{t,\ell}(u) = A_{t,\ell}(v) = A_{t,\ell}(w_\ell) = 1$  であるとする. ただし,  $k \neq \ell$  とする. このとき, ノード  $u$  からノード  $v$  への信頼リンク生成に対し,  $w_k$  と  $w_\ell$  が及ぼす影響は一般に異なると考えられる. したがって我々は, 媒介者の観点からアクティビティ共起の効果を分析するために, 任意の  $(u, v) \in C_{t+1}^*$  に対し確率  $P_{t+1}(u, v)$  を, A-ME モデルにアクティビティ共起情報を組み込むことにより,

$$P_{t+1}(u, v) = \sum_{k=1}^K \lambda_k \frac{1}{1 + \exp(-\boldsymbol{\theta}_k \cdot \mathbf{x}_{t,k}(u, v))}, \quad (2)$$

とモデル化することを提案する. ここで,  $\boldsymbol{\lambda}$  はパラメータベクトル

$$\boldsymbol{\lambda} = (\lambda_1, \dots, \lambda_K); \quad \sum_{k=1}^K \lambda_k = 1, \quad \lambda_k > 0 \quad (k = 1, \dots, K)$$

であり, 各  $\boldsymbol{\theta}_k \in \mathbf{R}^5$  はアクティビティ  $k$  に関するパラメータベクトル

$$\boldsymbol{\theta}_k = (\theta_{k,0}, \theta_{k,1}, \theta_{k,2}, \theta_{k,3}, \theta_{k,4})$$

である.  $\mathbf{x}_{t,k}(u, v)$  は時刻  $t_0 + t\Delta t$  におけるアクティビティ  $k$  に関する  $(u, v)$  の特徴ベクトル

$$\mathbf{x}_{t,k}(u, v) = (1, x_{t,k,1}(u, v), x_{t,k,2}(u, v), x_{t,k,3}(u, v), x_{t,k,4}(u, v))$$

であり,

$$x_{t,k,i}(u, v) = \sum_{w \in \mathcal{M}_{t,k,i}(u, v)} g(w) \quad (i = 1, 2, 3, 4)$$

である. ただし,  $\mathcal{M}_{t,k,i}(u, v)$  は, 期間  $I_t$  におけるタイプ  $i$  媒



介者のうち、ノード  $u$  およびノード  $v$  と期間  $I_t$  にアクティビティ  $k$  で共起した者全体の集合である。すなわち、

$$\mathcal{M}_{t,k,i}(u,v) = \{w \in \mathcal{M}_t(u,v) \mid A_{t,k}(u) = A_{t,k}(v) = A_{t,k}(w) = 1\}$$

である。 $\mathbf{x}_{t,k,i}(u,v)$  は、アクティビティ  $k$  で共起したという条件のもとで、ノード  $u$  からノード  $v$  へのタイプ  $i$  媒介者達全体の影響度を表していることに注意する。本論文では、信頼リンク生成ダイナミクスを媒介者アクティビティ共起の観点から分析するので、特に、任意の許容信頼リンク候補  $(u,v) \in C_{t+1}^*$  に対して、あるアクティビティで共起するような媒介者ノードが少なくとも1つ存在する、すなわち、あるアクティビティ  $k$  とタイプ  $i$  が存在し、

$$\mathcal{M}_{t,k,i}(u,v) \neq \emptyset$$

であると仮定する。我々は、媒介者タイプと媒介者アクティビティ共起の観点から信頼リンク生成ダイナミクスを分析する、このような確率モデルを A-MAE モデルと呼ぶ。

A-ME モデルは  $K=1$  の場合の A-MAE モデルとみなすことができることに注意する。A-MAE モデルおよび A-ME モデルのモデルパラメータの推定法については、ページ数の都合上ここでは説明を割愛する。

## 4. 実験

ソーシャルメディアでの信頼リンク生成に対し媒介者タイプ情報と媒介者アクティビティ共起情報の影響を調べるために、A-ME モデルと A-MAE モデルを用いて Epinions の実データを分析する。

### 4.1 ソーシャルメディアデータ

信頼リンク生成とユーザアクティビティが共に観測可能なソーシャルメディアとして、製品レビューサイト Epinions を対象とした。Epinions において、ユーザ  $u$  が他のユーザ  $v$  を Trust ユーザと登録したとき、本論文では、ユーザ  $u$  がユーザ  $v$  に信頼リンクを生成したとする。Epinions では、全ての製品は  $K$  個のカテゴリに分類されている。ユーザ  $u$  がカテゴリ  $k$  の製品にレビューもしくは評点を投稿したとき、本論文では、ユーザ  $u$  がアクティビティ  $k$  を実行したとする。

2012年10月に最人気ユーザと紹介されたユーザを起点として、新たなユーザが出現しなくなるまで、幅優先探索により信頼リンクを辿ることでユーザを探索し、各ユーザが生成した信頼リンクと投稿した製品レビューや評点情報を収集した。収集したデータは、ユーザ数 27,873 に対して信頼リンク数 218,686、製品数 268,897 に対してレビュー投稿数 809,521、評点投稿数 14,105,311 であった。また、カテゴリ数は  $K=19$  であった。

実験では、信頼リンクやアクティビティがコンスタントに生成もしくは実行された 2006 年と 2007 年のデータを用

いた。これを Epinions データと呼ぶ。

### 4.2 信頼リンク予測

媒介者が存在する 2 ユーザ間の近い将来の信頼リンク予測に関して A-ME モデルと A-MAE モデルを評価し、信頼リンク生成ダイナミクスのモデル化において、媒介者タイプ情報と媒介者アクティビティ共起情報を組み込むことの効果を調べる。観測データ系列  $\mathcal{D}_T$  に基づいてモデルパラメータ値を推定し、許容信頼リンク候補集合  $C_{T+1}^*$  に属する信頼リンクを予測することにより、それらの性能を評価する。以下では、実験で用いたベースライン法、媒介者価値スコアリング法、およびデータセットについて説明し、評価実験の結果を述べる。

#### 4.2.1 ベースラインと媒介者価値スコアリング

信頼リンク生成ダイナミクスのモデル化において、媒介者タイプ情報を組み込むことの効果を調べるために、媒介者の存在情報のみを用い、そのタイプ情報を考慮しないモデルをベースラインモデルとし、それと A-ME モデルをリンク予測性能で比較する。特に我々は、最も単純なベースラインモデルとして、A-ME モデルと同様、ロジスティック回帰モデルに基づき、任意の  $(u,v) \in C_{t+1}^*$  に対して確率  $P_{t+1}(u,v)$  を

$$P_{t+1}(u,v) = \frac{1}{1 + \exp(-\psi_0 - \psi_1 z_t(u,v))}$$

とモデル化するものを用いた。ここに、 $z_t(u,v)$  はノード  $u$  からノード  $v$  への媒介者全体の影響度を表し、

$$z_t(u,v) = \sum_{w \in \mathcal{M}_t(u,v)} g(w)$$

である。また、 $\boldsymbol{\psi} = (\psi_0, \psi_1) \in \mathbf{R}^2$  はパラメータベクトルである。

さて、媒介者  $w$  の媒介者価値のスコアリング  $g(w)$  に関しては、2つの代表的な手法 [1], [10] を実験で用いた。まず、すべての媒介者を同等に扱う、すなわち、

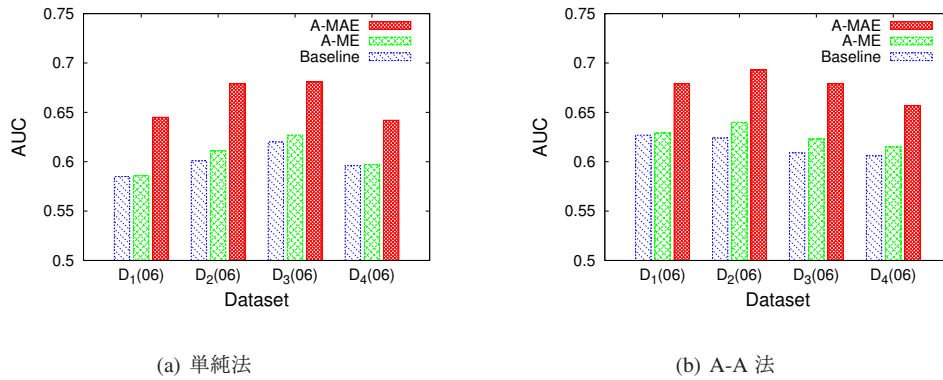
$$g(w) = 1$$

とする手法を調べた。このような媒介者価値スコアリング法を単純法と呼ぶ。単純法を用いるとき、任意の  $(u,v) \in C_{t+1}^*$  に対し  $z_t(u,v)$  は、期間  $I_t$  におけるノード  $u$  とノード  $v$  の共通近傍ノード数

$$z_t(u,v) = |\mathcal{N}_t(u) \cap \mathcal{N}_t(v)|$$

を表し、ベースラインモデルによる手法はリンク予測でよく用いられる共通近傍法 [1], [10] と同様なものになることに注意する。ただし本論文では、近傍ノードというときはリンクの向きを無視し、期間  $I_t$  におけるノード  $u'$  の近傍ノード全体の集合を

$$\mathcal{N}_t(u') = \{v' \in V \mid (u', v') \in E_t\} \cup \{v' \in V \mid (v', u') \in E_t\}$$

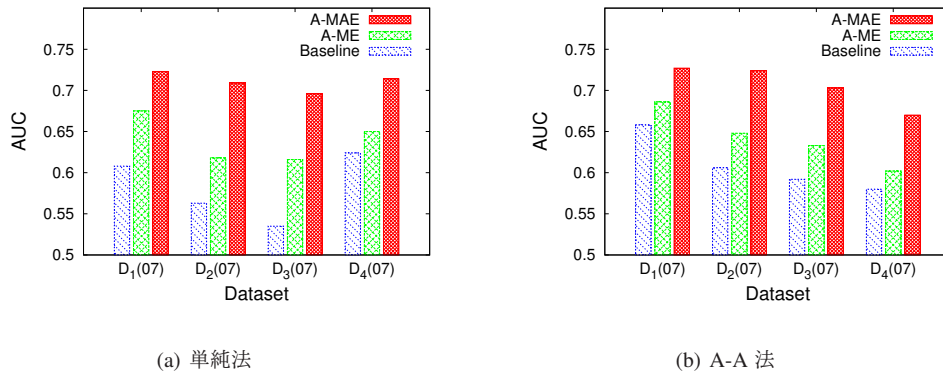


(a) 単純法

(b) A-A 法

図 2 2006 年 Epinions データにおける信頼リンク予測の性能比較

Fig. 2 Performance comparison of trust-link prediction in the 2006 Epinions data



(a) 単純法

(b) A-A 法

図 3 2007 年 Epinions データにおける信頼リンク予測の性能比較

Fig. 3 Performance comparison of trust-link prediction in the 2007 Epinions data

で表す. 次に, 媒介者  $w$  ごとにその価値  $g(w)$  が異なるようなスコアリング法として, 近傍ノード数が少ない媒介者ほどより大きな重みをおく, すなわち,

$$g(w) = 1/|N_r(w)|$$

とする手法を調べた. このスコアリング法を用いるとき, 任意の  $(u, v) \in C_{t+1}^*$  に対し  $z_t(u, v)$  は, 期間  $I_t$  におけるノード  $u$  とノード  $v$  の共通近傍の各ノードについて, その近傍ノード数の逆数を取り, それら全体で和をとったもの

$$z_t(u, v) = \sum_{w \in N_r(u) \cap N_r(v)} 1/|N_r(w)|$$

となるので, ベースラインモデルによる手法はリンク予測でよく用いられる Adamic-Adar 法 [1], [10] と同様なものになることに注意する. よって本論文では, このような媒介者価値スコアリング法を A-A 法と呼ぶ.

#### 4.2.2 データセット

実験では, 信頼ネットワーク進化のスピードとオンライン世界での時間経過に伴う情報の急速な陳腐化を考慮し, 最も単純なデータセットを複数構築するという観点から, 期間の長さ  $\Delta t$  を 3 ヶ月とし, 観測データ系列  $\mathcal{D}_T$  に関して  $T = 2$  とすることにより, Epinions の 2006 年データから 4 つのデータセット  $D_1(06), D_2(06), D_3(06), D_4(06)$  と, Epinions の 2007 年データから 4 つのデータセット  $D_1(07),$

$D_2(07), D_3(07), D_4(07)$  をそれぞれ構築した. ただし, 各データセットには, 初期時刻  $t_0$  および, 期間  $I_1, I_2, I_3$  における Epinions データが含まれている. このとき, 学習用のリンクデータは  $C_2^*$  であり,  $I_2$  内に生成された信頼リンクの集合  $C_2^* \cap E_2$  (正例) と  $I_2$  内に生成されなかった許容信頼リンク候補の集合  $C_2^* \setminus E_2$  (負例) から構成される. また, 評価用のリンクデータは  $C_3^*$  であり,  $I_3$  内に生成された信頼リンクの集合  $C_3^* \cap E_3$  (正例) と  $I_3$  内に生成されなかった許容信頼リンク候補の集合  $C_3^* \setminus E_3$  (負例) から構成される. ここで, 初期時刻  $t_0$  はそれぞれ,  $D_1(06)$  では 2006 年 1 月 1 日午前 0 時,  $D_2(06)$  では 2006 年 2 月 1 日午前 0 時,  $D_3(06)$  では 2006 年 3 月 1 日午前 0 時,  $D_4(06)$  では 2006 年 4 月 1 日午前 0 時であり,  $D_1(07), D_2(07), D_3(07), D_4(07)$  についても同様である. 例えば,  $D_1(06)$  では, 期間  $I_1$  は 2006 年 1 月 1 日から 2006 年 3 月 31 日, 期間  $I_2$  は 2006 年 4 月 1 日から 2006 年 6 月 30 日, 期間  $I_3$  は 2006 年 7 月 1 日から 2006 年 9 月 30 日である.

#### 4.2.3 評価結果

構築した 8 個のデータセットに対し, 2 つの媒介者価値スコアリング法を用いて, A-MAE モデル, A-ME モデル, およびベースラインモデルの信頼リンク予測性能を比較した. 任意の許容信頼リンク候補  $(u, v) \in C_3^*$  に対して,  $(u, v)$  が期間  $I_3$  に信頼リンクとして生成される確率  $P_3(u, v)$  を

考え、各モデルによって得られる確率  $P_3(u, v)$  の推定値を  $\hat{P}_3(u, v)$  とする。各モデルは、 $\hat{P}_3(u, v)$  の値にしたがってリンク  $(u, v) \in C_3^*$  をランキングすることにより、信頼リンク予測を行う。本論文では、予測精度は ROC 曲線に対する AUC で測定した。

図 2 と図 3 にそれぞれ、2006 年 Epinions データと 2007 年 Epinions データに対する結果を示す。ここに、媒介者価値スコアリングに単純法を用いた結果を図 2(a) と図 3(a) に、媒介者価値スコアリングに A-A 法を用いた結果を図 2(b) と図 3(b) に示している。単純法と A-A 法のどちらの媒介者価値スコアリング法を用いても、すべてのデータセットにおいて、A-ME モデルはベースラインモデルよりも精度が高く、さらに、A-MAE モデルは A-ME モデルよりも大幅に精度が向上していることが観察される。したがって、信頼リンク生成ダイナミクスのモデル化においては、媒介者タイプ情報を組み込むことは効果があり、さらに、媒介者アクティビティ共起情報とそれに基づく媒介者タイプ情報を組み込むことは非常に有効であると言える。すなわち、信頼リンク生成における媒介者の影響に関しては、共起するアクティビティが異なればその影響の度合いは異なり、また、同じアクティビティで共起しても、媒介者タイプが異なればその影響の度合いは異なることがわかる。

## 5. まとめ

ソーシャルメディアにおける信頼ネットワークの進化過程を説明できるモデルの構築を目指して、本論文では、信頼リンク生成において媒介者が果たす役割を分析することに焦点をあてた。そのために、媒介者が存在する 2 ユーザ間の信頼リンク生成ダイナミクスを分析するモデルとして、2 つの確率過程モデル A-ME と A-MAE を提案した。ここに、A-ME モデルは信頼ネットワーク進化における媒介者タイプ情報の効果を分析するものであり、また A-MAE モデルは、A-ME モデルの拡張であって、さらに媒介者アクティビティ共起情報の効果を分析するものである。我々は A-ME モデルと A-MAE モデルを用いて Epinions データを分析した。媒介者価値のスコアリングに関しては、すべての媒介者を同等に扱う単純法と媒介者をその信頼リンクの数で価値づけする A-A 法という、2 つの代表的な手法を実験で用いた。まず、媒介者が存在する 2 ユーザ間における近い将来の信頼リンク予測に関して、A-ME モデルと A-MAE モデルを評価した。A-ME モデルを媒介者価値スコアリングの仕方に対応したベースライン法と比較することにより、媒介者タイプ情報を組み込むことは、信頼リンク生成ダイナミクスのモデル化において効果があることを実証した。また、どちらの媒介者価値スコアリング法を用いても、A-ME モデルに比べて A-MAE モデルは信頼リンク予測精度において大幅な性能向上を実現できることを示し、媒介者アクティビティ共起情報とそれに基づく媒介者

タイプ情報を組み込むことは、信頼リンク生成ダイナミクスのモデル化において非常に有効であることを実証した。

謝辞 本研究は JSPS 科研費 26330352 の助成を受けたものである。

## 参考文献

- [1] Barbieri, N., Bonchi, F. and Manco, G.: Who to follow and why: link prediction with explanations, *Proceedings of KDD 2014*, pp. 1266–1275 (2014).
- [2] Chen, W., Lakshmanan, L. and Castillo, C.: Information and influence propagation in social networks, *Synthesis Lectures on Data Management*, Vol. 5, pp. 1–177 (2013).
- [3] Crandall, D., Cosley, D., Huttenlocher, D., Kleinberg, J. and Suri, S.: Feedback effects between similarity and social influence in online communities, *Proceedings of KDD 2008*, pp. 160–168 (2008).
- [4] Gomez-Rodriguez, M., Leskovec, J. and Krause, A.: Inferring Networks of Diffusion and Influence, *Proceedings of KDD 2010*, pp. 1019–1028 (2010).
- [5] Guha, R., Kumar, R., Raghavan, P. and Tomkins, A.: Propagation of trust and distrust, *Proceedings of WWW 2004*, pp. 403–412 (2004).
- [6] Kempe, D., Kleinberg, J. and Tardos, E.: Maximizing the spread of influence through a social network, *Proceedings of KDD 2003*, pp. 137–146 (2003).
- [7] Kimura, M., Saito, K., Nakano, R. and Motoda, H.: Extracting influential nodes on a social network for information diffusion, *Data Mining and Knowledge Discovery*, Vol. 20, pp. 70–97 (2010).
- [8] Kimura, M., Saito, K., Ohara, K. and Motoda, H.: Learning to predict opinion share and detect anti-majority opinionists in social networks, *Journal of Intelligent Information Systems*, Vol. 41, pp. 5–37 (2013).
- [9] Leskovec, J., Huttenlocher, D. and Kleinberg, J.: Predicting positive and negative links in online social networks, *Proceedings of WWW 2010*, pp. 641–650 (2010).
- [10] Liben-Nowell, D. and Kleinberg, J.: The link-prediction problem for social networks, *Journal of the American Society for Information Science and Technology*, Vol. 58, pp. 1019–1031 (2007).
- [11] Liu, H., Lim, E., Lauw, H., Le, M., Sun, A., Srivastava, J. and Kim, Y.: Predicting trusts among users of online communities: an opinion case study, *Proceedings of EC 2008*, pp. 310–319 (2008).
- [12] Mannila, H. and Terzi, E.: Finding links and initiators: a graph-reconstruction problem, *Proceedings of SDM 2009*, pp. 1207–1217 (2009).
- [13] Nguyen, V., Lim, E., Jiang, J. and Sun, A.: To trust or not to trust? predicting online trusts using trust antecedent framework, *Proceedings of ICDM 2009*, pp. 896–901 (2009).
- [14] Saito, K., Kimura, M., Ohara, K. and Motoda, H.: Learning asynchronous-time information diffusion models and its application to behavioral data analysis over social networks, *Journal of Computer Engineering and Informatics*, Vol. 1, pp. 30–57 (2013).
- [15] Tang, J., Gao, H., Hu, X. and Liu, H.: Exploiting homophily effect for trust prediction, *Proceedings of WSDM 2013*, pp. 53–62 (2013).
- [16] Tang, J., Gao, H., Liu, H. and Sarma, A. D.: eTrust: understanding trust evolution in an online world, *Proceedings of KDD 2012*, pp. 253–261 (2012).