

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¹ 熊野 雅仁² 木村 昌弘² 齊藤 和巳³ 大原 剛三⁴ 元田 浩⁵

受付日 2014年11月10日, 再受付日 2014年12月22日/2015年4月17日,
採録日 2015年4月17日

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, 自然な確率過程モデル *A-ME* を考え, その拡張である *A-MAE* モデルを提案する. 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-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. 我々は, 提案モデルに対して, モデルパラメータの値を信頼リンクとユーザアクティビティの観測系列データから推定する効率的な手法を与える. そして, 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. さらに, *A-ME* モデルと *A-MAE* モデルを用いて, Epinions での信頼リンク生成に関するいくつかの特徴的な性質を明らかにする.

キーワード: 信頼ネットワーク進化分析, 媒介者タイプ, 媒介者アクティビティ共起

Analysis of Mediator-activity Effects for Trust-network Evolution in Social Media

KEITO HATTA¹ MASAHITO KUMANO² MASAHIRO KIMURA² KAZUMI SAITO³ KOUZOU OHARA⁴
HIROSHI MOTODA⁵

Received: November 10, 2014, Revised: December 22, 2014/April 17, 2015,
Accepted: April 17, 2015

Abstract: We analyze evolution of trust networks in social media sites from a perspective of mediators. To this end, we consider the *A-ME* model as a natural stochastic model that analyzes the dynamics of creating a trust link under the presence of mediators, and propose the *A-MAE* model that is an extension of the *A-ME* model. Here, the *A-ME* model analyzes mediator-type effects for trust-network evolution in terms of mediator types, and the *A-MAE* model analyzes mediator-activity effects for trust-network evolution. We present an efficient method of inferring the values of model parameters from an observed sequence of trust links and user activities. Using real data from Epinions, we experimentally show that the *A-ME* model outperforms the baseline method and the *A-MAE* model significantly outperforms the *A-ME* model for predicting trust links in the near future under the presence of mediators, and demonstrate the effectiveness of mediator-type and mediator-activity information for trust-network evolution. We further clarify, by using the *A-ME* and *A-MAE* models, several characteristic properties of trust-link creation in the Epinions data.

Keywords: trust-network evolution, mediator type, mediator activity

¹ 龍谷大学大学院理工学研究科電子情報学専攻
Division of Electronics and Informatics, Ryukoku University,
Otsu, Shiga 520-2194, Japan

² 龍谷大学理工学部電子情報学科
Department of Electronics and Informatics, Ryukoku Uni-
versity, Otsu, Shiga 520-2194, Japan

³ 静岡県立大学経営情報学部経営情報学科
School of Administration and Informatics, University of
Shizuoka, Shizuoka 422-8526, Japan

⁴ 青山学院大学理工学部情報テクノロジー学科
Department of Integrated Information Technology, Aoyama
Gakuin University, Sagami-hara, Kanagawa 229-8558, Japan

⁵ 大阪大学産業科学研究研究所
Institute of Scientific and Industrial Research, Osaka Uni-
versity, Ibaraki, Osaka 567-0047, Japan

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も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], [7], [8], [9], [16]。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た [4], [5], [11], [12], [13], [15], [17], [18]。

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 [10] 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。オンライン世界での情報はしばしば時間経過とともに急速に陳腐化していくので、我々は特に、最近形成された媒介者のみに焦点をあて、それら媒介者が近い将来の信頼リンク生成に果たす役割を調べる。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。たとえば、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り、一方、タイプ 2 媒介者 w' とは、 u から w' への信頼リンクと v から w' 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うると、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 1 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、ユーザたちは、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、

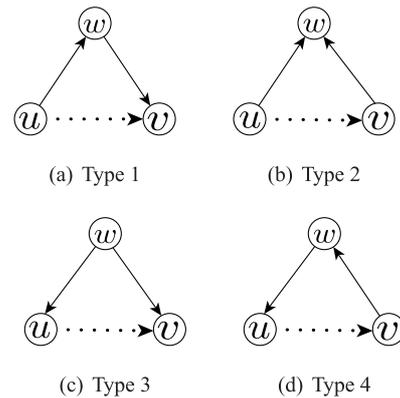


図 1 ノード u からノード v への媒介者 w の 4 タイプ
Fig. 1 Four types of mediator w from node u to node v .

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。

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、自然な確率過程モデル $A-ME$ (*Analyzing Mediator Effects*) を考え、その拡張である $A-MAE$ (*Analyzing Mediator-Activity Effects*) モデルを提案する。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、 $A-MAE$ 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。我々は、 $A-ME$ モデルと $A-MAE$ モデルに対して、モデルパラメータの値を信頼リンクとユーザアクティビティの観測系列データから効率良く推定する手法を与える。そして、 $A-ME$ 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^{*1}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

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

非常に有効であることを実証する．さらに，A-ME モデルと A-MAE モデルを用いて，Epinions での信頼リンク生成に関するいくつかの特徴的な性質を明らかにする．

本論文*2の構成は以下のとおりである．2章では，論文で用いる記号法を述べる．3章で A-ME モデルと提案する A-MAE モデルを定義し，4章でそれらのモデルにおけるパラメータ推定法を与える．5章では，Epinions の実データを用いた実験および分析の結果を報告する．さらに，6章は，関連研究を述べる．そして，7章でまとめる．

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ーが行う，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ー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． Δt 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． $\bar{G}_t = (V, \bigcup_{s=1}^t E_s)$ は時刻 $t_0 + t\Delta t$ でのユーザー集合 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

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

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．したがって， 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，ユーザーは 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，ユーザー 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 $A_{t,k}(u) = 1$

*2 本論文は，国際会議 PRICAI'14 において研究の基本アイデアと初期結果を発表したもの [6] の拡張であり，これを関連研究に照らし合わせて系統的かつ詳細に論じるとともに，媒介者価値のスコアリングとそれに基づく信頼リンク予測のベースライン法を導入し，広範かつ詳細な評価実験を行った結果を述べたものである．

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 $\mathcal{M}_t(u, v)$ とする． $\mathcal{M}_t(u, v) = \mathcal{M}_t(v, u)$ 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 $\mathcal{M}_{t,1}(u, v)$ 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 $\mathcal{M}_{t,2}(u, v)$ 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 $\mathcal{M}_{t,3}(u, v)$ とし， $(w, u), (v, w) \in E_t$ ならば w をタイプ 4 媒介者と呼び (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， $P_{t+1}(u, v)$ の自然なモデル A-ME (Analyzing Mediator Effects) を考え，その拡張であ

る *A-MAE (Analyzing Mediator-Activity Effects)* モデルを提案する．ここに、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、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*3.

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

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

とモデル化することを考える．ここで、 $\phi \in \mathbf{R}^5$ はパラメータベクトル

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

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

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

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_ℓ 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_ℓ は期間 I_t においてアクティビティ ℓ を実行（アクティビティ ℓ 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_ℓ 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、

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

とモデル化することを提案する．ここで、 λ はパラメータベクトル

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

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

$$\theta_k = (\theta_{k,0}, \theta_{k,1}, \theta_{k,2}, \theta_{k,3}, \theta_{k,4})$$

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、

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

である． $x_{t,k,i}(u, v)$ 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．

*3 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 (5.2.1 項参照)．

4. 推定法

T を 2 以上の整数とし, $t = 1, \dots, T$ に対して, 期間 I_t における信頼リンク集合 E_t とユーザアクティビティ集合 $\mathcal{A}_t = \{\mathbf{A}_t(u) \mid u \in V\}$ の観測データが与えられているとし, それら観測系列を \mathcal{D}_T とする. すなわち,

$$\mathcal{D}_T = \{(E_t, \mathcal{A}_t) \mid t = 1, \dots, T\}$$

とする. 本章では, A-ME モデルおよび A-MAE モデルに対して, モデルパラメータの値を観測系列 \mathcal{D}_T から推定する手法を与える. A-ME モデルは $K = 1$ の場合の A-MAE モデルと見なすことができるので (式 (1), (2) 参照), 以下では A-MAE モデルについての推定法のみを述べる.

MAP 推定の枠組みに従ってモデルパラメータの値の推定を行う. λ に対してディリクレ事前分布を, $\theta_1, \dots, \theta_K$ に対してガウス事前分布を, それぞれ仮定する. そして, 目的関数

$$\begin{aligned} \mathcal{L}(\lambda, \theta_1, \dots, \theta_K) &= \sum_{t=1}^{T-1} \left(\sum_{(u,v) \in C_{t+1}^* \cap E_{t+1}} \log \left(\sum_{k=1}^K \frac{\lambda_k}{1 + \exp(-\theta_k \cdot \mathbf{x}_{t,k}(u,v))} \right) \right. \\ &+ \sum_{(u,v) \in C_{t+1}^* \setminus E_{t+1}} \log \left(\sum_{k=1}^K \frac{\lambda_k \exp(-\theta_k \cdot \mathbf{x}_{t,k}(u,v))}{1 + \exp(-\theta_k \cdot \mathbf{x}_{t,k}(u,v))} \right) \left. \right) \\ &+ \gamma \sum_{k=1}^K \log \lambda_k - \sum_{k=1}^K \frac{1}{2\sigma_k^2} \sum_{i=0}^4 \theta_{k,i}^2, \end{aligned} \quad (3)$$

を最大にする λ と $\theta_1, \dots, \theta_K$ の値を推定する (式 (2) 参照). ここで, $\gamma > 0$ と $\sigma_k > 0$ ($k = 1, \dots, K$) は正則化定数である. 我々は, EM アルゴリズムによりパラメータの最適値を求める. $\bar{\lambda} = (\bar{\lambda}_1, \dots, \bar{\lambda}_K)$ を λ の現在の推定値とし, $k = 1, \dots, K$ に対して $\bar{\theta}_k = (\bar{\theta}_{k,0}, \dots, \bar{\theta}_{k,4})$ を θ_k の現在の推定値とする. 式 (3) から Jensen の不等式より,

$$\begin{aligned} \mathcal{L}(\lambda, \theta_1, \dots, \theta_K) - \mathcal{L}(\bar{\lambda}, \bar{\theta}_1, \dots, \bar{\theta}_K) &\geq Q(\lambda, \theta_1, \dots, \theta_K \mid \bar{\lambda}, \bar{\theta}_1, \dots, \bar{\theta}_K), \end{aligned} \quad (4)$$

を得る. ここに,

$$\begin{aligned} Q(\lambda, \theta_1, \dots, \theta_K \mid \bar{\lambda}, \bar{\theta}_1, \dots, \bar{\theta}_K) &= \sum_{t=1}^{T-1} \sum_{k=1}^K \left(\sum_{(u,v) \in C_{t+1}^* \cap E_{t+1}} \bar{c}_{t,k}(u,v) \{ \log \lambda_k \right. \\ &- \log (1 + \exp(-\theta_k \cdot \mathbf{x}_{t,k}(u,v))) \} \\ &+ \sum_{(u,v) \in C_{t+1}^* \setminus E_{t+1}} \bar{d}_{t,k}(u,v) \{ \log \lambda_k - \theta_k \cdot \mathbf{x}_{t,k}(u,v) \\ &- \log (1 + \exp(-\theta_k \cdot \mathbf{x}_{t,k}(u,v))) \} \left. \right) + \gamma \sum_{k=1}^K \log \lambda_k \\ &- \sum_{k=1}^K \frac{1}{2\sigma_k^2} \sum_{i=0}^4 \theta_{k,i}^2 + \bar{Q}_0(\bar{\lambda}, \bar{\theta}_1, \dots, \bar{\theta}_K), \end{aligned} \quad (5)$$

であり, また, 各 (t, k, u, v) に対して,

$$\bar{c}_{t,k}(u,v) = \frac{\bar{\lambda}_k}{1 + \exp(-\bar{\theta}_k \cdot \mathbf{x}_{t,k}(u,v))} \quad (6)$$

$$\sum_{\ell=1}^K \frac{\bar{\lambda}_\ell}{1 + \exp(-\bar{\theta}_\ell \cdot \mathbf{x}_{t,\ell}(u,v))}$$

$$\bar{d}_{t,k}(u,v) = \frac{\bar{\lambda}_k \exp(-\bar{\theta}_k \cdot \mathbf{x}_{t,k}(u,v))}{1 + \exp(-\bar{\theta}_k \cdot \mathbf{x}_{t,k}(u,v))} \quad (7)$$

$$\sum_{\ell=1}^K \frac{\bar{\lambda}_\ell \exp(-\bar{\theta}_\ell \cdot \mathbf{x}_{t,\ell}(u,v))}{1 + \exp(-\bar{\theta}_\ell \cdot \mathbf{x}_{t,\ell}(u,v))}$$

である. ただし, $\bar{Q}_0(\bar{\lambda}, \bar{\theta}_1, \dots, \bar{\theta}_K)$ は, $Q(\bar{\lambda}, \bar{\theta}_1, \dots, \bar{\theta}_K \mid \bar{\lambda}, \bar{\theta}_1, \dots, \bar{\theta}_K) = 0$ となる $\bar{\lambda}$ と $\bar{\theta}_1, \dots, \bar{\theta}_K$ の関数である. 式 (4), (5) から, λ と $\theta_1, \dots, \theta_K$ の更新式は, λ と $\theta_1, \dots, \theta_K$ に関して $Q(\lambda, \theta_1, \dots, \theta_K \mid \bar{\lambda}, \bar{\theta}_1, \dots, \bar{\theta}_K)$ を最大化することにより導出することができる. したがって, 式 (5), (6), (7) から, まず λ についての更新式

$$\lambda_k = \frac{\sum_{t=1}^{T-1} \left(\sum_{(u,v) \in C_{t+1}^* \cap E_{t+1}} \bar{c}_{t,k}(u,v) + \sum_{(u,v) \in C_{t+1}^* \setminus E_{t+1}} \bar{d}_{t,k}(u,v) \right) + \gamma}{\sum_{\ell=1}^K \sum_{t=1}^{T-1} \left(\sum_{(u,v) \in C_{t+1}^* \cap E_{t+1}} \bar{c}_{t,\ell}(u,v) + \sum_{(u,v) \in C_{t+1}^* \setminus E_{t+1}} \bar{d}_{t,\ell}(u,v) \right) + \gamma K}$$

($k = 1, \dots, K$) を得ることができる. 次に, $\theta_1, \dots, \theta_K$ に関する更新式を導出する. 式 (5), (6), (7) から, $k = 1, \dots, K$ と $i = 0, 1, 2, 3, 4$ に対して,

$$\begin{aligned} \frac{\partial Q}{\partial \theta_{k,i}} &= \sum_{t=1}^{T-1} \left(\sum_{(u,v) \in C_{t+1}^* \cap E_{t+1}} \frac{\bar{c}_{t,k}(u,v) x_{t,k,i}(u,v) \exp(-\theta_k \cdot \mathbf{x}_{t,k}(u,v))}{1 + \exp(-\theta_k \cdot \mathbf{x}_{t,k}(u,v))} \right. \\ &- \sum_{(u,v) \in C_{t+1}^* \setminus E_{t+1}} \left. \frac{\bar{d}_{t,k}(u,v) x_{t,k,i}(u,v)}{1 + \exp(-\theta_k \cdot \mathbf{x}_{t,k}(u,v))} \right) - \frac{1}{\sigma_k^2} \theta_{k,i} \end{aligned} \quad (8)$$

を得る. さらに, $k = 1, \dots, K$ と $i, j = 0, 1, 2, 3, 4$ に対して,

$$\begin{aligned} \frac{\partial^2 Q}{\partial \theta_{k,i} \partial \theta_{k,j}} &= - \sum_{t=1}^{T-1} \left(\sum_{(u,v) \in C_{t+1}^* \cap E_{t+1}} \frac{\alpha_{t,k,i,j}(u,v)}{\{1 + \exp(-\theta_k \cdot \mathbf{x}_{t,k}(u,v))\}^2} \right. \\ &+ \sum_{(u,v) \in C_{t+1}^* \setminus E_{t+1}} \left. \frac{\beta_{t,k,i,j}(u,v)}{\{1 + \exp(-\theta_k \cdot \mathbf{x}_{t,k}(u,v))\}^2} \right) \\ &- \frac{1}{\sigma_k^2} \delta_{i,j} \end{aligned} \quad (9)$$

を得る. ただし,

$$\begin{aligned} & \alpha_{t,k,i,j}(u,v) \\ &= \bar{c}_{t,k}(u,v)x_{t,k,i}(u,v)x_{t,k,j}(u,v)\exp(-\theta_k \cdot x_{t,k}(u,v)) \\ & \beta_{t,k,i,j}(u,v) \\ &= \bar{d}_{t,k}(u,v)x_{t,k,i}(u,v)x_{t,k,j}(u,v)\exp(-\theta_k \cdot x_{t,k}(u,v)) \end{aligned}$$

であり、 $\delta_{i,j}$ はクロネッカーのデルタである。式 (9) から、行列 $(\partial^2 Q / \partial \theta_{k,i} \partial \theta_{k,j})$ は負定値であることが保証される。よって、方程式 $\partial Q / \partial \theta_{k,i} = 0$ ($k = 1, \dots, K, i = 0, 1, 2, 3, 4$) (式 (8) 参照) を解くことにより、 $Q(\lambda, \theta_1, \dots, \theta_K | \bar{\lambda}, \bar{\theta}_1, \dots, \bar{\theta}_K)$ を最大化する $\theta_1, \dots, \theta_K$ の値を求めることができる。すなわち、 $\theta_1, \dots, \theta_K$ についての更新式が得られる。実験では、標準的なニュートン法を用いて (式 (8), (9) 参照)、それらの更新式を計算した。

5. 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。まず、近い将来の信頼リンク予測に関して、A-ME モデルと A-MAE モデルの性能を評価する。次に、A-ME モデルと A-MAE モデルに基づいて、Epinions での信頼リンク生成に関するユーザー行動の性質を分析する。

5.1 ソーシャルメディアデータ

信頼リンク生成とユーザーアクティビティがとも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、ユーザー u が他のユーザー v を Trust ユーザと登録したとき、本論文では、ユーザー u がユーザー v に信頼リンクを生成したとする。Epinions では、すべての製品は K 個のカテゴリに分類されている。ユーザー 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、ユーザー u がアクティビティ k を実行したとする。

2012 年 10 月に最人気ユーザーと紹介されたユーザーを起点として、新たなユーザーが出現しなくなるまで、幅優先探索に

より信頼リンクをたどることでユーザーを探索し、各ユーザー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、ユーザー数 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 $K = 19$ であり、次のようなものであった。

- 1) Hotels & Travel, 2) Wellness & Beauty, 3) Restaurants & Gourmet, 4) Web Sites & Internet Services, 5) Media,
- 6) Home & Garden, 7) Kids & Family, 8) Business & Technology, 9) Computers & Internet, 10) Electronics,
- 11) Games, 12) Sports & Outdoors, 13) Personal Finance, 14) Gifts, 15) Cars & Motorsports, 16) Education, 17) Local Services,
- 18) Miscellaneous, 19) Archived Reviews.

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。図 2(a) に 1 カ月内に生成された信頼リンク数の変動を示し、図 2(b) に 1 カ月内に実行されたアクティビティ数の変動を示す。次に、Epinions データに対して、その基本的な統計的性質を調べた。信頼ネットワークの入次数分布と出次数分布 [14] をそれぞれ、図 3(a) と図 3(b) に示す。これらはべき乗則 (power-law) [14] に従うことが観察される。アクティビティ数の分布、つまり、各ユーザーが実行したアクティビティ数 (頻度) の分布を図 3(c) に示す。これもべき乗則 [14] に従っていることが観察される。したがって、Epinions データは、ソーシャルネットワークやユーザーアクティビティのデータが示す典型的な統計的性質を持つといえる。

5.2 信頼リンク予測

媒介者が存在する 2 ユーザー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 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

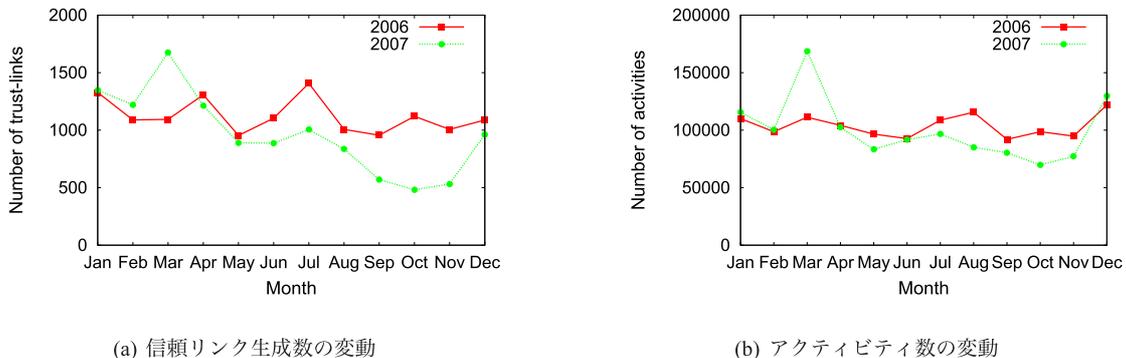


図 2 Epinions データの変動
Fig. 2 Fluctuation in the Epinions data.

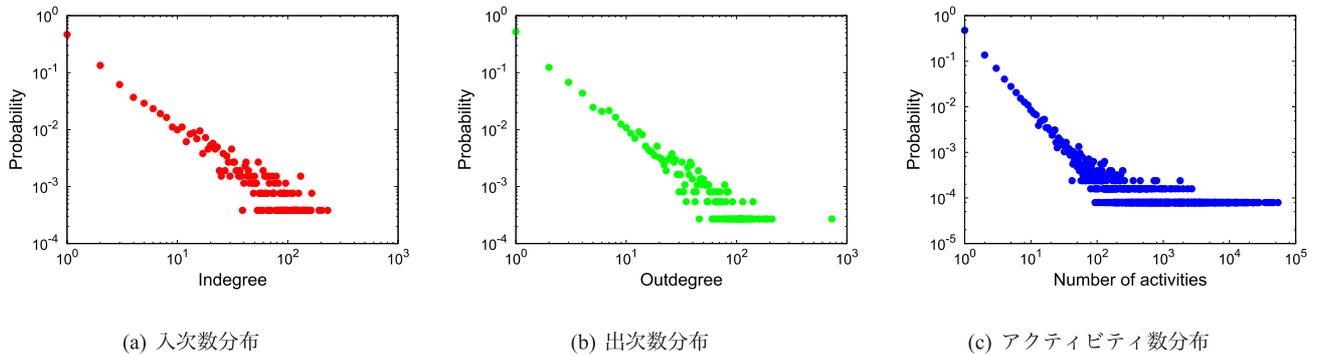


図 3 Epinions データの基本統計解析

Fig. 3 Fundamental statistical analysis of the Epinions data.

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。

5.2.1 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、 $\psi = (\psi_0, \psi_1) \in \mathbf{R}^2$ 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], [11] 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], [11] 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

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

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、

$$g(w) = 1/|\mathcal{N}_t(w)|$$

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

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

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], [11] 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。

5.2.2 データセットと実験設定

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ともな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 Δt を 3 カ月とし、観測データ系列 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. たとえば, $D_1(06)$ 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. 各データセットにおける学習用リンク数と評価用リンク数を表 1 に示す. 学習および評価にはある程度十分なデータ量があり, また, 一般的なリンク予測問題と同様, 正例と負例の不均衡は見られるが, 信頼リンク候補全体にしろる正例の割合は, 媒介者アクティビティ共起が存在する場合 (すなわち, C_2^* と C_3^*) に限定したことにより, そのように限定しない一般の場合と比べて 1,000 倍以上高くなっていることに注意する.

次に, 各データセットにおいて, 各タイプの媒介者がどのくらい存在し, その中での重複はどのくらいであったかについて述べる. ここに, ノード u からノード v への媒介者 w は一般に複数のタイプの媒介者でありうるが, 媒介者タイプの重複において可能なものは, 「①タイプ 1 とタイプ 2, ②タイプ 1 とタイプ 3, ③タイプ 2 とタイプ 4, ④タイプ 3 とタイプ 4, ⑤タイプ 1 からタイプ 4 までの 4 タイプすべて」の 5 パターンに限られることに注意する

表 1 データセットの基本統計量
Table 1 Fundamental statistics of datasets.

Dataset	学習用リンクデータ		評価用リンクデータ	
	$ C_2^* \cap E_2 $	$ C_2^* \setminus E_2 $	$ C_3^* \cap E_3 $	$ C_3^* \setminus E_3 $
$D_1(06)$	673	60967	376	18915
$D_2(06)$	672	53270	393	17860
$D_3(06)$	710	49668	391	19151
$D_4(06)$	613	47130	306	17166
$D_1(07)$	903	94747	282	22081
$D_2(07)$	734	88024	171	13676
$D_3(07)$	564	75956	163	10086
$D_4(07)$	538	47182	284	18383

(図 1 参照). 各データセットにおけるノード u からノード v への媒介者に対して, 各タイプおよび各重複パターンの媒介者数の平均値を表 2 に示す. すべてのデータセットで同様な傾向が見られた. また, 紙数の都合から詳細な統計を示すのを省略したが, 各データセットにおいて学習用データの媒介者統計と評価用データの媒介者統計は, ほぼ同じであったことを追記しておく. 表 2 より, タイプ 3 媒介者が最も多くタイプ 1 媒介者が最も少なかったこと, また, 重複パターンに関しては, タイプ 3 でもありタイプ 4 でもある媒介者が最も多く, 次いでタイプ 1 でもありタイプ 3 でもある媒介者が多かったことが分かる. さらに, 媒介者タイプ別に他のタイプとの重複の割合を調べた. タイプ 1 媒介者とタイプ 4 媒介者は, とともに約 72% が重複して他のタイプでもあった. タイプ 2 媒介者の約 58% は重複して他のタイプでもあった. 最も重複が少なかったのはタイプ 3 媒介者であり, その割合は約 53% であった. 特に, 「あるタイプの媒介者のほぼすべてが, 他のタイプの媒介者であるということとはなかった」ということに注意する.

モデルパラメータ推定におけるハイパーパラメータ γ , σ_k ($k = 1, \dots, K$) については, 標準的な設定である, $\gamma = 1$ (ラプラススムージング) と $\sigma_k = 1$ ($k = 1, \dots, K$) (θ_k の事前分布として平均 0 で分散 1 のガウス分布) を用いた. また, モデルパラメータ推定における終了条件については, 目的関数 (式 (3) 参照) の値の増加率が十分小さくなったとき, すなわち,

$$\frac{\mathcal{L}(\lambda, \theta_1, \dots, \theta_K) - \mathcal{L}(\bar{\lambda}, \bar{\theta}_1, \dots, \bar{\theta}_K)}{|\mathcal{L}(\bar{\lambda}, \bar{\theta}_1, \dots, \bar{\theta}_K)|} < 10^{-4}$$

とした. ここに, $\bar{\lambda}, \bar{\theta}_1, \dots, \bar{\theta}_K$ はそれぞれ, パラメータベクトル $\lambda, \theta_1, \dots, \theta_K$ の更新前のものを表している (4 章参照). A-ME モデルの場合は, ニュートン法による $\theta_1, \dots, \theta_K$ のみの更新となるが, 実験では 5, 6 回程度のパラメータ更新で収束していた. A-MAE モデルの場合は, λ の更新と $\theta_1, \dots, \theta_K$ の更新を合わせて, 30 回程度のパラメータ更新で収束していた.

5.2.3 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

表 2 データセットの媒介者統計
Table 2 Mediator statistics of datasets.

	$D_1(06)$	$D_2(06)$	$D_3(06)$	$D_4(06)$	$D_1(07)$	$D_2(07)$	$D_3(07)$	$D_4(07)$
$ \mathcal{M}_{t,1}(u, v) $	0.403	0.415	0.418	0.429	0.384	0.420	0.416	0.429
$ \mathcal{M}_{t,2}(u, v) $	0.457	0.473	0.493	0.498	0.419	0.447	0.454	0.421
$ \mathcal{M}_{t,3}(u, v) $	0.659	0.618	0.606	0.608	0.754	0.722	0.683	0.744
$ \mathcal{M}_{t,4}(u, v) $	0.424	0.435	0.437	0.445	0.405	0.442	0.438	0.456
$ \mathcal{M}_{t,1}(u, v) \cap \mathcal{M}_{t,2}(u, v) $	0.171	0.174	0.179	0.189	0.169	0.190	0.182	0.184
$ \mathcal{M}_{t,1}(u, v) \cap \mathcal{M}_{t,3}(u, v) $	0.213	0.206	0.203	0.219	0.214	0.236	0.224	0.254
$ \mathcal{M}_{t,2}(u, v) \cap \mathcal{M}_{t,4}(u, v) $	0.174	0.177	0.181	0.190	0.172	0.193	0.185	0.186
$ \mathcal{M}_{t,3}(u, v) \cap \mathcal{M}_{t,4}(u, v) $	0.222	0.213	0.211	0.225	0.223	0.247	0.235	0.267
$ \bigcap_{i=1}^4 \mathcal{M}_{t,i}(u, v) $	0.092	0.088	0.089	0.100	0.097	0.112	0.101	0.111

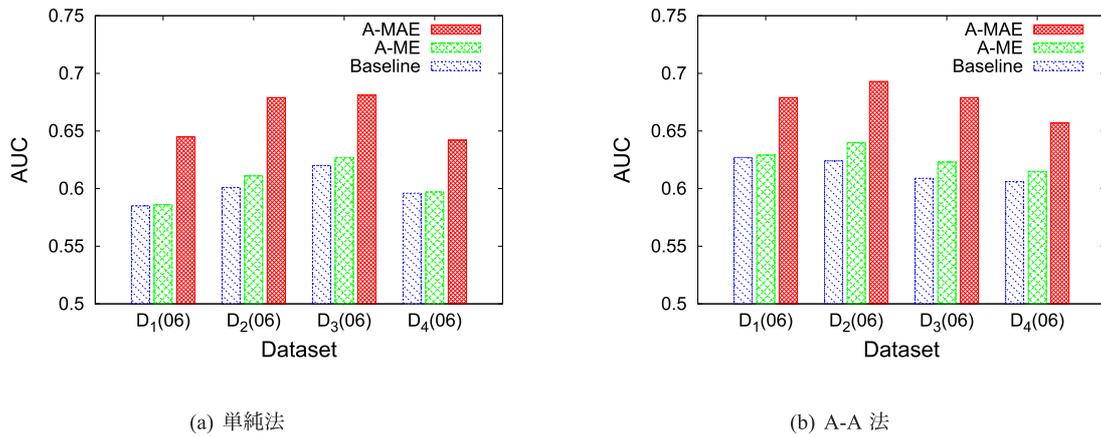


図 4 2006 年 Epinions データにおける信頼リンク予測の性能比較

Fig. 4 Performance comparison of trust-link prediction in the 2006 Epinions data.

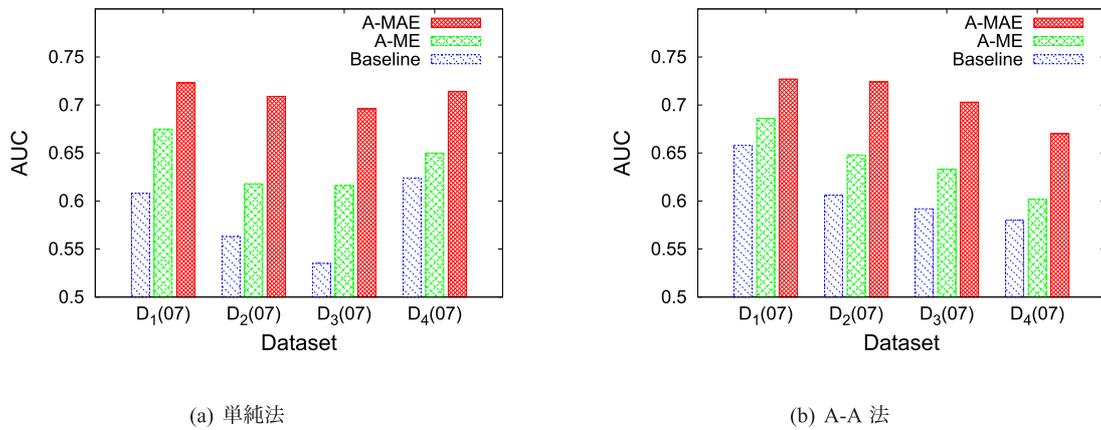


図 5 2007 年 Epinions データにおける信頼リンク予測の性能比較

Fig. 5 Performance comparison of trust-link prediction in the 2007 Epinions data.

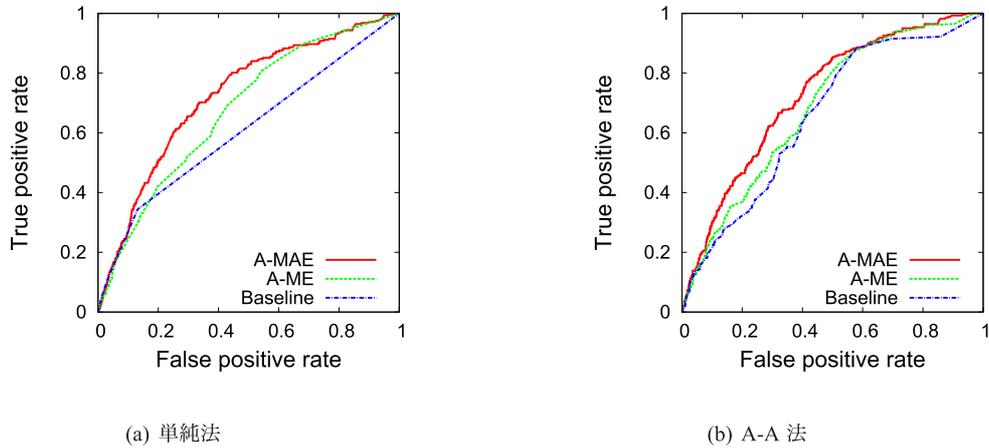


図 6 データセット $D_1(07)$ における ROC 曲線

Fig. 6 ROC curves for dataset $D_1(07)$.

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、 $\hat{P}_3(u, v)$ の値に従ってリンク $(u, v) \in C_3^*$ 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 4 と図 5 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 4(a) と図 5(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 4(b) と図 5(b) に示している。また、ROC 曲線に関する

る結果の例として、図 6 にデータセット $D_1(07)$ での結果を示す。図 6(a) は単純法を、そして図 6(b) は A-A 法を、それぞれ媒介者価値スコアリングに用いた場合の結果である。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いえる。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分かる。ところで、A-MAE モデルの予測性能を考えると、媒介者価値スコアリングに関しては、ほぼすべてのデータセットにおいて A-A 法が単純法をやや上回ったが、そうではない場合も見られた。媒介者価値スコアリング法の検討は重要であるが、それについては今後の課題としたい。

5.3 影響分析

A-ME モデルと A-MAE モデルに基づいて、Epinions データでの信頼リンク生成に関するユーザ行動の性質を調

べた。ここでは、データセット $D_1(07)$ の学習期間 I_1, I_2 における分析結果のみを述べる。すなわち、2007 年 1 月から 2007 年 6 月までの期間における Epinions データを分析する。媒介者価値スコアリングに関しては、予測精度の高かった A-A 法を用いる。

図 7 に、A-MAE モデルにおけるパラメータ $\lambda = (\lambda_k)$ の推定値を示す。共起するアクティビティ k が異なれば、信頼リンク生成における媒介者の影響 λ_k は異なることが見てとれる。この期間において影響力が強い上位 5 つのアクティビティは、 $k = 1$ (Hotels & Travel), $k = 2$ (Wellness & Beauty), $k = 3$ (Restaurants & Gourmet), $k = 4$ (Web Sites & Internet Services), $k = 5$ (Media) であった。

次に、アクティビティ k ($k = 1, 2, 3, 4, 5$) に対して A-

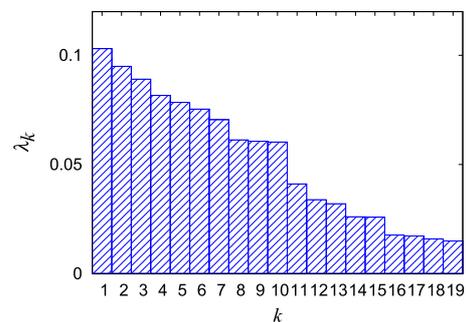


図 7 データセット $D_1(07)$ における $\lambda = (\lambda_k)$ の推定値 (A-A 法)
 Fig. 7 Estimated values of $\lambda = (\lambda_k)$ for dataset $D_1(07)$ (A-A method).

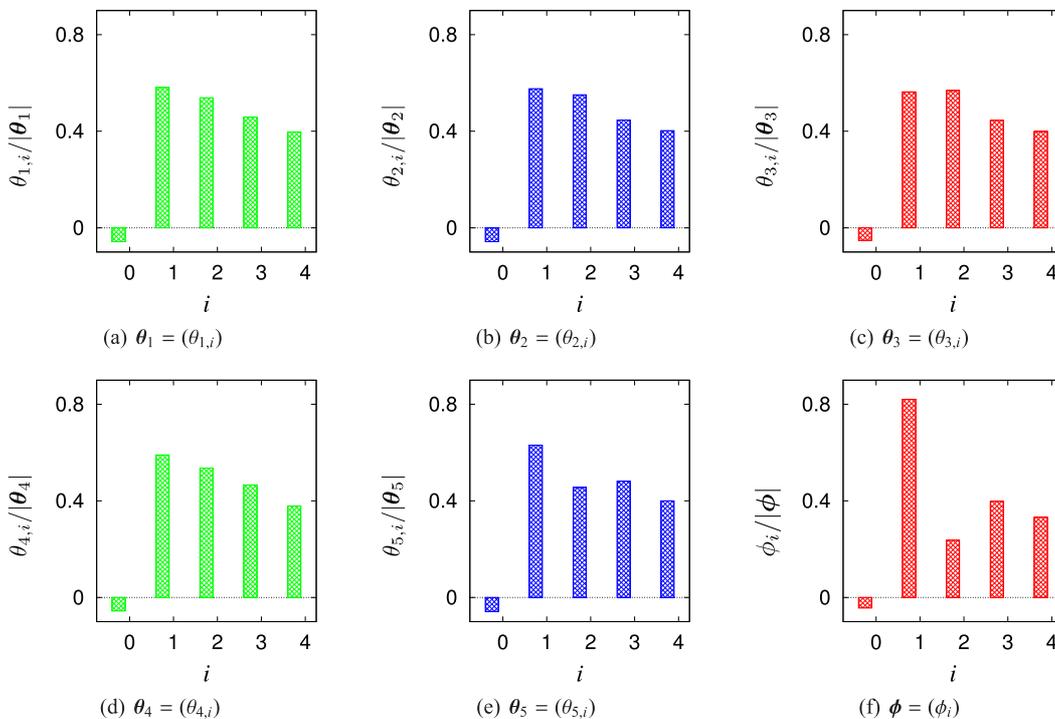


図 8 データセット $D_1(07)$ における θ_k ($k = 1, 2, 3, 4, 5$) (A-MAE モデル) と ϕ (A-ME モデル) の推定値 (A-A 法)

Fig. 8 Estimated values of θ_k ($k = 1, 2, 3, 4, 5$) (A-MAE model) and ϕ (A-ME model) for dataset $D_1(07)$ (A-A method).

MAE モデルのパラメータ θ_k と、媒介者アクティビティ共起情報を用いない A-ME モデルのパラメータ ϕ の推定値を調べた。ここに、 ϕ は θ_k ($k = 1, \dots, 19$) のある種の平均と見なされることに注意する。図 8 に θ_k ($k = 1, 2, 3, 4, 5$)、 ϕ の推定値を示す。ただし、それらはユークリッドノルムが 1 となるように正規化している。まず、各アクティビティ k に対して、媒介者タイプ i が異なれば^{*4}、信頼リンク生成における媒介者の影響 $\theta_{k,i}$ は異なることが確認できる。また、 ϕ では、タイプ 1 媒介者の影響度 ϕ_1 がその他のタイプの媒介者の影響度に比べて突出して高く、タイプ 2 媒介者の影響度 ϕ_2 が最も低い。上位 5 つのアクティビティ k でも、タイプ 1 媒介者の影響度 $\theta_{k,1}$ は高い。しかしながら、上位 4 つのアクティビティ k では、タイプ 2 媒介者の影響度 $\theta_{k,2}$ も同様に高いことが観察される。特に、アクティビティ $k = 3$ では、タイプ 2 媒介者の影響度 $\theta_{3,2}$ がその他のタイプの媒介者の影響度に比べて最も高い。したがって、この期間の Epinions の信頼リンク生成では、タイプ 2 媒介者は、上位 4 つのアクティビティにおいて特に強い影響力を持っていたという特徴的な性質があることが分かる。

6. 関連研究

媒介者タイプに着目したリンク推定に関しては、1 章において述べたように Leskovec らの研究 [10] がある。彼らは信頼リンクのようなポジティブリンクと不信リンクのようなネガティブリンクが混在するソーシャルネットワークデータに対して、社会心理学におけるバランス理論とステータス理論を拡張し、トライアド構造 (媒介者タイプ) に基づくロジスティック回帰モデルを構築してリンク推定問題を研究した。そして、与えられたリンクの符号を推定する問題およびポジティブリンクの存在を推定する問題において、トライアド構造やネガティブリンク情報の有効性を示している。しかしながら、彼らの研究では時間情報を考慮しておらず、本論文で対象としているようなネットワーク進化ダイナミクスを扱うことを想定していない。ところで、ネガティブリンク情報を用いずにポジティブリンクを予測することを考える場合に、もし彼らの手法に自然に時間情報を組み込むならば、それは媒介者価値スコアリングに単純法を利用した A-ME モデルになることに注意しておく。近年、ネガティブリンク情報を持つソーシャルネットワークデータの収集が困難になりつつあり、本論文ではそ

のような情報を持たないネットワークデータを主対象としたが、我々の提案モデルを符号つきネットワークに拡張することは容易である。符号つき信頼ネットワーク進化における媒介者アクティビティ共起効果の分析は、今後の重要課題の 1 つである。

ところで、信頼リンクとアクティビティ類似性の相互関係に着目して、信頼リンクの存在を推定する研究もなされている。Tang ら [17] は、信頼ネットワークの隣接行列に対する低ランク行列分解モデルに、評点づけ行動におけるユーザ達のアクティビティ類似性を組み込む hTrust を提案し、製品レビューサイトデータに対してその有効性を示した。Barbieri ら [1] は、リンクとアクティビティを同時に生成する確率的潜在変数モデルとして WTFW モデルを提案し、リンクの存在を推定することおよび、リンクがトピック的なものか社会的なものかを説明することにおいて有効性を示した。これらの研究も時間情報を考慮していないのでネットワーク進化ダイナミクスを直接扱うことはできないが、そこで提案されたモデルを我々の提案モデルと融合させることで、より精緻なネットワーク進化ダイナミクスモデルを構築できる可能性がありうる。そのような方向性を探求することも、今後の課題である。

上記のように、オンライン世界の信頼ネットワークに関するほとんどの研究はユーザ間の静的な信頼関係を仮定しており、信頼ネットワーク進化を直接扱った研究はあまりない。オンライン世界の信頼ネットワーク進化に関するものとしては、製品レビューサイトに焦点をあてた Tang らの研究 [18] が代表的である。彼らの研究では、現在までのユーザの評点づけ行動と信頼ネットワークの成長過程に基づき、近い将来にユーザがつける評点および生成される信頼リンクを予測する枠組みとして、協調フィルタリングでよく用いられる潜在因子分析モデルを拡張した eTrust を提案し、実データを用いた実験で有効性を示した。ところで eTrust では、ユーザの評点づけ行動のモデル化が主対象であり、信頼リンク生成ダイナミクスは直接モデル化されていない。本論文では、信頼リンク形成における最も基本的な分析の 1 つとして、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したが、eTrust では、それに関して媒介者タイプやアクティビティ共起の効果を直接分析することはできない。しかしながら eTrust は、評点情報や時間減衰効果など提案モデルでまだ考慮していない特徴量をも組み込んでおり、提案モデルの拡張の方向性を示唆している。提案モデルを拡張し、信頼リンク予測性能において eTrust と比較することは今後の課題である。

7. 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

^{*4} タイプ 1 媒介者の影響度が高いということは、信頼している人が信頼している人に信頼リンクを生成しやすい傾向があることを意味し、タイプ 2 媒介者の影響度が高いということは、信頼している人が同じである人に信頼リンクを生成しやすい傾向があることを意味し、タイプ 3 媒介者の影響度が高いということは、信頼してくれている人が同じである人に信頼リンクを生成しやすい傾向があることを意味し、タイプ 4 媒介者の影響度が高いということは、信頼してくれている人を信頼している人に信頼リンクを生成しやすい傾向があることを意味する。

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、自然な確率過程モデル A-ME を考え、その拡張である A-MAE モデルを提案した。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、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 モデルに対して、モデルパラメータの値を信頼リンクとユーザアクティビティの観測系列データから効率良く推定する手法を与え、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。また、異なる媒介者アクティビティや異なる媒介者タイプは、ともに信頼リンク生成に与える影響も異なるということを実証した。さらに、A-ME モデルと A-MAE モデルを用いて、媒介者アクティビティ共起と媒介者タイプの観点から Epinions での信頼リンク生成に関するいくつかの特徴的な性質を明らかにした。

本論文では、Epinions に関する分析結果を詳細に述べたが、提案法を他のソーシャルメディアの信頼ネットワークデータに適用し、その有効性を検証することは緊急の課題である。また、媒介者の価値のスコアリングについては代表的な 2 手法を調べたが、その他の手法に関する検討は重要であり、今後の課題と考えている。現実の現象を十分説明できるように提案モデルを高度化するにはまだ多くの課題はあるが、本研究において提案モデルの有望性が示されたことは、ソーシャルメディアにおける信頼ネットワーク進化のモデル構築に対して、重要な第一歩であると我々は考えている。

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, *Proc. 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD'14)*, 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, *Proc. 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD'08)*, pp.160–168 (2008).
- [4] Gomez-Rodriguez, M., Leskovec, J. and Krause, A.: Inferring Networks of Diffusion and Influence, *Proc. 16th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD'10)*, pp.1019–1028 (2010).
- [5] Guha, R., Kumar, R., Raghavan, P. and Tomkins, A.: Propagation of trust and distrust, *Proc. 13th International Conference on World Wide Web (WWW'04)*, pp.403–412 (2004).
- [6] Hatta, K., Kumano, M., Kimura, M., Saito, K., Ohara, K. and Motoda, H.: Analyzing mediator-activity effects for trust-network evolution in social media, *Proc. 13th Pacific Rim International Conference on Artificial Intelligence (PRICAI'14)*, LNAI, Vol.8862, pp.297–308, Springer (2014).
- [7] Kempe, D., Kleinberg, J. and Tardos, E.: Maximizing the spread of influence through a social network, *Proc. 9th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD'03)*, pp.137–146 (2003).
- [8] 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).
- [9] 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).
- [10] Leskovec, J., Huttenlocher, D. and Kleinberg, J.: Predicting positive and negative links in online social networks, *Proc. 19th International Conference on World Wide Web (WWW'10)*, pp.641–650 (2010).
- [11] 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).
- [12] 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, *Proc. 9th ACM Conference on Electronic Commerce (EC'08)*, pp.310–319 (2008).
- [13] Mannila, H. and Terzi, E.: Finding links and initiators: A graph-reconstruction problem, *Proc. 2009 SIAM International Conference on Data Mining (SDM'09)*, pp.1207–1217 (2009).
- [14] Newman, M.E.J.: The structure and function of complex networks, *SIAM Review*, Vol.45, pp.167–256 (2003).
- [15] Nguyen, V., Lim, E., Jiang, J. and Sun, A.: To trust or not to trust? Predicting online trusts using trust antecedent framework, *Proc. 9th IEEE International Conference on Data Mining (ICDM'09)*, pp.896–901 (2009).

- [16] 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).
- [17] Tang, J., Gao, H., Hu, X. and Liu, H.: Exploiting homophily effect for trust prediction, *Proc. 6th ACM International Conference on Web Search and Data Mining (WSDM'13)*, pp.53-62 (2013).
- [18] Tang, J., Gao, H., Liu, H. and Sarma, A.D.: eTrust: Understanding trust evolution in an online world, *Proc. 18th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD'12)*, pp.253-261 (2012).



八田 圭斗

1990年生。2013年龍谷大学工学部電子情報学科卒業。2015年同大学大学院工学研究科電子情報学専攻修士課程修了。同年株式会社日立ソリューションズ・クリエイト入社。



熊野 雅仁 (正会員)

1991年立命館大学工学部基礎工学部卒業。1991年龍谷大学工学部実験助手。2008年同大学工学部実験講師。現在に至る。映像理論、コンテンツ制作支援技術、ソーシャルメディアネットワーク等に関する研究と教育に従事。博士(工学)(神戸大学大学院工学研究科情報知能学専攻)。電子情報通信学会、人工知能学会、日本音響学会、認知科学会、映像情報メディア学会、芸術科学会、日本ヴァーチャルリアリティ学会、IEEE、ACM各会員。



木村 昌弘

1987年大阪大学理学部数学科卒業。1989年同大学大学院理学研究科数学専攻修士課程修了。同年日本電信電話株式会社(NTT)入社。NTTコミュニケーション科学基礎研究所等を経て、2005年退社。現在、龍谷大学工学部電子情報学科教授。複雑ネットワーク科学、データマイニング、機械学習等の研究と教育に従事。博士(理学)。日本数学会、日本応用数理学会、人工知能学会、電子情報通信学会各会員。



齊藤 和巳 (正会員)

1963年生。1985年慶應義塾大学工学部数理科学科卒業。同年NTT入社。1991年より1年間オタワ大学客員研究員。2007年静岡県立大学経営情報学部教授。機械学習、複雑ネットワーク等の研究に従事。博士(工学)(東京大学)。電子情報通信学会、人工知能学会、日本神経回路学会、日本応用数理学会、日本データベース学会、日本行動計量学会、観光情報学会各会員。



大原 剛三 (正会員)

1995年大阪大学大学院基礎工学研究科博士前期課程修了。1996年日本学術振興会特別研究員DC2。1997年大阪大学産業科学研究所助手、同助教を経て、2009年より青山学院大学工学部情報テクノロジー学科准教授。データマイニング、機械学習、社会ネットワーク分析の研究に従事。博士(工学)。IEEE、AAAI、電子情報通信学会、人工知能学会各会員。



元田 浩

1965年東京大学工学部原子力工学科卒業。1967年同大学大学院原子力工学専攻修士課程修了。同年日立製作所に入社。同社中央研究所、原子力研究所、エネルギー研究所、基礎研究所を経て、1995年退社。1996年大阪大学産業科学研究所教授(知能システム科学研究部門、高次推論研究分野)、2006年定年退職し、現在、米国空軍科学技術局東京オフィス(AFOSR/AOARD)科学顧問。大阪大学名誉教授、同招聘教授、タズマニア大学客員教授。原子力システムの設計、運用、診断、制御に関する研究を経て、機械学習、知識獲得、知識発見、データマイニング、社会ネットワーク解析の研究に従事。工学博士。