

機関番号：34316

研究種目：基盤研究(C)

研究期間：2008～2010

課題番号：20500147

研究課題名(和文) 複雑ネットワーク上での連続時間情報拡散過程の数理モデリング

研究課題名(英文) Modeling continuous-time information diffusion process
on complex networks

研究代表者

木村 昌弘(KIMURA MASAHIRO)

龍谷大学・理工学部・教授

研究者番号：10396153

研究成果の概要(和文): 社会ネットワークは、イノベーション、ホットトピック、さらに悪意のある噂など、様々な情報を拡散させる媒体として重要な役割を果たす。本研究では、複雑ネットワーク上での情報拡散現象に対し、機械学習アプローチを用いてその予測可能な数理モデルを構築し、それに基づいて、与えられた制約の下で情報拡散を助長および抑制する手法の考案や、情報拡散におけるノードの振る舞いの分析法の考案など、各種応用法を考案し評価した。

研究成果の概要(英文): A social network can play an important role as a medium for the spread of various types of information, including innovation, hot topics and even malicious rumors. In this research, we have constructed predictable models for information diffusion phenomena in complex networks by exploiting a machine learning approach. Moreover, we have proposed and evaluated such a variety of applications of the models that present the methods of increasing and decreasing the information spread under given constraints, and analyzing behaviors of nodes in information propagation.

交付決定額

(金額単位：円)

	直接経費	間接経費	合計
2008年度	1,200,000	360,000	1,560,000
2009年度	1,100,000	330,000	1,430,000
2010年度	1,100,000	330,000	1,430,000
年度			
年度			
総計	3,400,000	1,020,000	4,420,000

研究分野： 知能情報学、情報数理学

科研費の分科・細目：情報学・知能情報学

キーワード： 社会ネットワーク分析、情報拡散モデル、機械学習、影響最大化、汚染最小化、行動データ分析、データマイニング、複雑ネットワーク科学

1. 研究開始当初の背景

(1) 人間社会や自然界における多くのシステムは、活動単位をノード、活動関係をリンクで表すとネットワークとして表現することが可能である。コンピュータなどの発達により大規模ネットワークを収集し分析できるようになってきたため、近年、複雑ネットワーク(complex network)研究が注目されている。複雑ネットワークの例とし

ては、インターネット、World Wide Web、ブログ、企業のビジネス関係、知人関係、電力網、神経網および蛋白質や遺伝子の相互作用関係などがあげられる。複雑ネットワーク研究では、ネットワークで表現されるシステムの動的な振る舞い(機能)を、ネットワークの構造的特徴に基づいて理解し予測することが重要な課題となっている。特に、社会ネットワーク上での情報、アイ

デア、イノベーションおよび影響などの拡散現象や、コンピュータネットワーク上でのコンピュータウイルスの拡散現象など、複雑ネットワーク上での拡散現象の理解や予測は、重要な研究課題の一つである。

(2) 先行研究では、Kempe、Kleinberg および Tardos が、社会ネットワーク上での離散時間の情報拡散過程モデルの基本確率モデルである、Independent Cascade (IC) モデルと Linear Threshold (LT) モデルに対し、影響最大化問題を近似的に解くアルゴリズムを与えていた。ここに、影響最大化問題とは、社会ネットワーク、その上の情報(影響) 拡散モデルおよびターゲットノード数が与えられたとき、その社会ネットワーク上で情報(影響) が広がるノード数の期待値を最大にするにはどのノード集合をターゲットノード集合(最初に情報を伝えるまたは影響を与えるノード集合)とするかという組み合わせ最適化問題であり、バイラルマーケティング(viral marketing)と関係している。我々は、Kempe らによる影響最大化問題の解法の効率化を実現する手法を提案していた。

2. 研究の目的

(1) 既存研究におけるネットワーク上の情報拡散モデルは、一つのノードが複数のノードに情報を伝えることができないという制約や、離散時間モデルであるなど、現実の情報拡散現象を十分な精度で説明するには限界があった。本研究では、上記のような制約をもたない、ネットワーク上での(パラメータをもつ)連続時間情報拡散過程の数理モデルを考案し、機械学習アプローチにより、複雑ネットワーク上での情報拡散現象を十分な精度で説明できる予測可能な数理モデルを構築する。

(2) 構築した情報拡散モデルに基づいて、与えられた制約の下で情報拡散を助長および抑制する手法の考案や、社会ネットワーク上での情報拡散におけるノードの振る舞いの分析法(行動分析法)の考案など、情報拡散モデルの各種応用法を考案する。

3. 研究の方法

(1) 現実のプログロールネットワーク上での情報拡散過程を調査して、その統計的性質を明らかにし、ネットワーク上での情報拡散の数理モデル構築のために利用する。さらに、プログロールネットワークにおける情報拡散のデータセットを作成し、本研究で構築する数理モデルの評価用データとして用いる。

(2) 社会ネットワーク上での離散時間の情報

拡散過程モデルの基本確率モデルである、IC モデルとLT モデルを拡張することにより、ネットワーク上の情報拡散過程を十分な精度で説明しうる、(パラメータをもつ)連続時間情報拡散過程モデルを構築する。また、本モデルのパラメータ推定法を考案し、上記(1)で作成した情報拡散のデータセットで評価する。そして、精度が上がらなければ、モデルの改良やパラメータ推定法の改良を行う。

(3) ICおよびLTモデルに対する影響最大化問題に対して我々が先に提案していた手法を拡張して、構築した情報拡散モデルに対する影響最大化問題や汚染最小化問題の効率的な解法を考案し、実ネットワークポロジを用いた数値実験で評価する。また、ネットワーク上での情報拡散過程を分析するために、構築した情報拡散モデルに基づいたネットワーク可視化法を考案し、実ネットワークポロジを用いた実験で評価する。さらに、構築した情報拡散モデルに基づいて、情報拡散におけるノードの振る舞いを分析する手法を考案し、上記(1)で作成した情報拡散のデータセットで評価する。その他、情報拡散モデルの各種応用法を考案し評価する。

4. 研究成果

(1) ネットワーク上の連続時間情報拡散過程の数理モデルとして、離散時間の基本確率モデルである IC モデルと LT モデルに、連続時間の非同期時間遅れを組み込んだ、AsIC モデルと AsLT モデルを提案した。これらモデルでは、決定する必要がある複数のパラメータが存在する。AsIC モデルでは時間遅れパラメータ r と拡散パラメータ p であり、AsLT モデルでは時間遅れパラメータ r と重みパラメータ q である。ネットワーク上での情報拡散の観測データからこれらパラメータの値を推定する手法を提案した。実ネットワークポロジを用いた実験により、提案法は限られた観測データから真の拡散モデルを高精度に同定できることを実証した。

1 まず、提案学習法がモデルパラメータを高精度に推定できることを、実ネットワークポロジを用いた実験で実証した。プログネットワークにおいて、観測データ数(アクティブノード数)に対する提案法の AsIC モデルと AsLT モデルのパラメータ推定誤差率を、それぞれ表 1 と表 2 に示す。

表 1 AsIC モデルに対する推定結果

#{active nodes}	Estimation error for r	Estimation error for p
1,163	0.019	0.026
5,151	0.018	0.014
10,322	0.011	0.011

表 2 AsLT モデルに対する推定結果

#{active nodes}	Estimation error for r	Estimation error for q
1,023	0.020	0.020
5,018	0.012	0.020
10,037	0.012	0.020

次に、提案法により学習したモデルが真のモデルにおける強影響ノードを高精度に

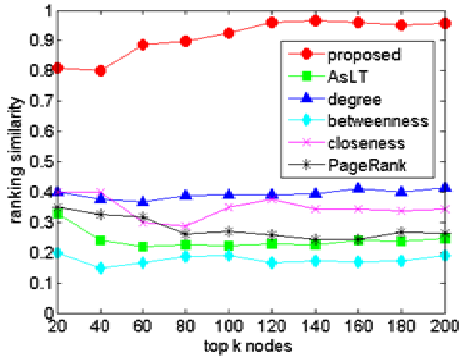


図 1 AsICモデルに対する抽出結果

抽出できることを、実ネットワークポロジを用いた実験で実証した。プログネットワークにおける AsIC モデルの強影響ノード群抽出に対して、社会ネットワーク分析の標準的な手法と提案法との性能比較結果を、図 1 に示す。同様に、AsLT モデルに対する結果を図 2 に示す。

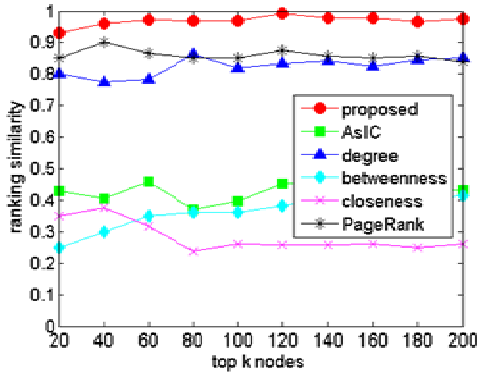


図 2 AsLTモデルに対する抽出結果

(2) 社会ネットワークにおける情報拡散過程を、AsIC モデルと AsLT モデルを用いて分析する手法を提案し評価した。

1 まず、観測データから推定した拡散モデルのパラメータのペアを 2 次元空間にプロットすることにより、トピック伝搬における人々の振る舞いを分析した。AsIC モデルに基づいた提案分析法を、現実のプログネットワークにおける情報拡散データに適用した結

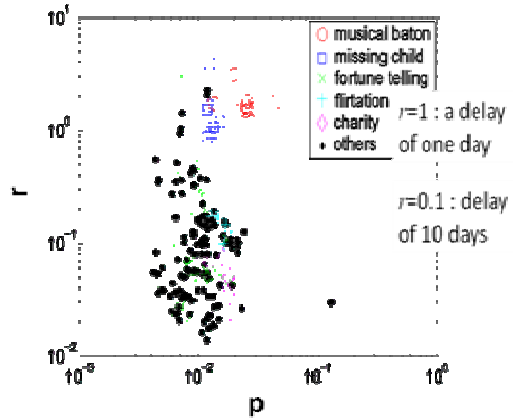


図 3 AsICモデルに基づいた分析結果

果を図 3 に示す。例えば、提案分析法では、"missing child"トピックが、高い拡散確率で速く伝搬している、緊急性のあるトピックであることを検出している。

次に、観測データをより良く説明するモデルを選択する手法として、KL ダイバージェンスに基づいた新たな手法を構築し、実ネットワークポロジを用いた実験でその有

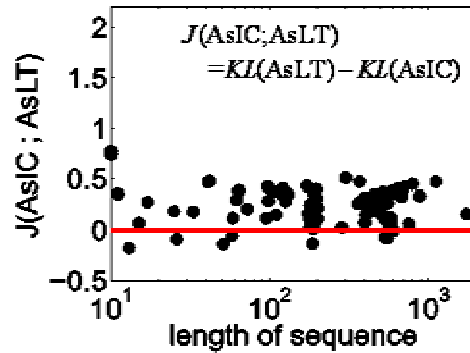


図 4 AsICモデル選択の結果

効性を実証した。プログネットワークにおいて、AsIC モデルから生成された観測系列の長さモデル選択の精度の関係を図 4 に示す。

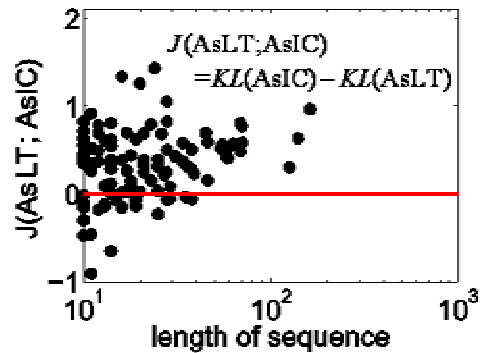


図 5 AsLTモデル選択の結果

同様に AsLT モデルに対する結果を図 5 に示す。現実のプログネットワークにおける情報

拡散データに適用した結果を図6に示す。

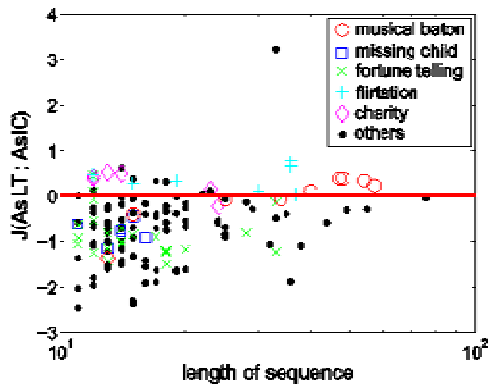


図6 モデル選択の結果

(3) 代表的な SIR 型離散時間モデルの IC モデルと LT モデルにおける影響最大化問題に対して、我々が提案していた手法の有効性を、数理的に説明するとともに、実ネットワークポロジを用いた実験において詳細に評価することにより明らかにした。

(4) ネットワークにおいて指定された数のリンク群を封鎖することにより好ましくない情報の広がりを最小化するという、影響最

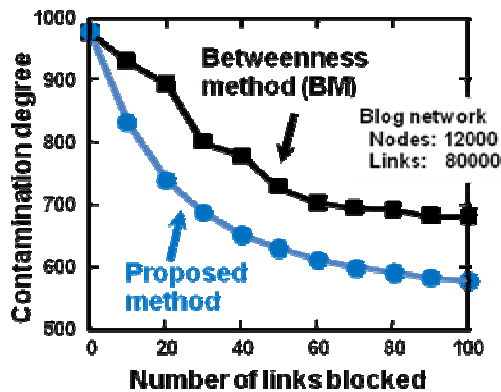


図7 ICモデルに対する汚染最小化問題の解の高品質化

大化問題と反対の問題である汚染最小化問題を、ICおよびLTモデルに対して考察した。我々は、汚染最小化問題の近似解を自然な貪欲戦略に基づいて効率よく求める新たな手法を提案し、その有効性を数理的に説明するとともに、大規模な実ネットワークを用いた実験により提案法が既存法よりも高品質であることを実証した。ブログネットワークでの IC モデルにおける、従来法 (betweenness 法) と提案法の性能比較結果を、図7に示す。

(5) SIR 型モデルでは、ノードは一度アクティブになると、再び非アクティブにはならないと仮定されているが、社会ネットワークにおける情報拡散現象には、非アクティブに戻

って再度アクティブになる SIS 型モデルとして考えるべき例も多々ある。そこで、IC モデルを SIS 型に拡張した離散時間の情報拡散モデルに対して、ノードの影響度の効率的推定法および影響最大化問題の効率的解法を構築し、それらの有効性を大規模な実ネットワークを用いた実験により実証した。

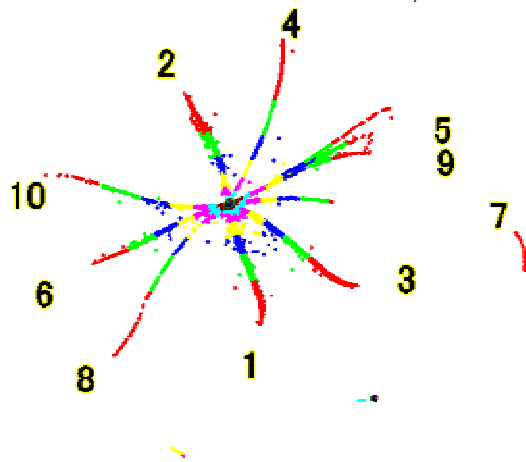


図8 ICモデルに基づく可視化

(6) 複雑ネットワークを情報拡散過程という動的な角度から理解する可視化法を提案し、大規模な実ネットワークデータを用いた実験により、提案法が既存法よりも直観と合致する可視化結果が生成可能であることを実証した。IC モデルと LT モデルに基づいたブログネットワークの可視化結果を、それぞれ図8と図9に示す。

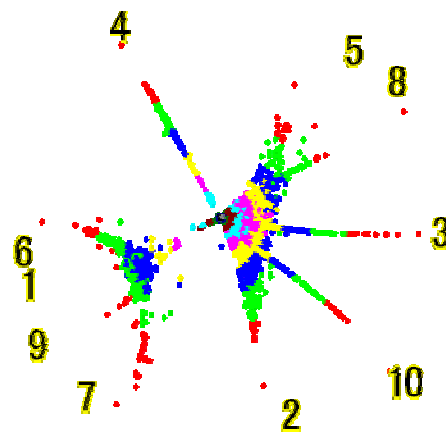


図9 LTモデルに基づく可視化

(7) IC モデルを SIS 型に拡張した非同期時間遅れをもつ情報拡散モデルに対して、ノードの影響度を効率よく計算する手法を構築した。また、情報拡散における super-mediator を発見する手法を構築した。さらに、価値重みつき投票者モデルを用いて、異なる価値を

もつ複数オピニオンの拡散データから、各オピニオンの価値を推定し、その将来シェアを予測する手法を構築し、その有効性を実証した。また、文書ストリームの主要潜在トピック抽出法など、情報拡散モデルの実問題への応用を提案し、実データを用いて有効性を実証した。

5. 主な発表論文等

(研究代表者、研究分担者及び連携研究者には下線)

[雑誌論文](計25件)

K.Saito, M.Kimura, K.Ohara, and H.Motoda, Efficient discovery of influential nodes for SIS models in social networks, Knowledge and Information Systems, 査読有, to appear

M.Kimura, K.Saito, K.Ohara, and H.Motoda, Learning information diffusion model in a social network for predicting influence of nodes, Intelligent Data Analysis, 査読有, Vol.15 (4), 2011, in press

K.Saito, M.Kimura, K.Ohara, and H.Motoda, Detecting changes in opinion value distribution for voter model, Lecture Notes in Computer Science (Proc. of SBP 2011), 査読有, Vol.6589, 2011, 89-96

M.Kimura, K.Saito, K.Ohara, and H.Motoda, Learning to predict opinion share in social networks, Proceedings of the Twenty-Fourth AAAI Conference on Artificial Intelligence (AAAI-10), 査読有, 2010, 1364-1370

K.Saito, M.Kimura, K.Ohara, and H.Motoda, Generative models of information diffusion with asynchronous timedelay, JMLR Workshop Conference Proceedings (Proc. of ACML 2010), 査読有, Vol.13, 2010, 193-208

K.Saito, M.Kimura, K.Ohara, and H.Motoda, Discovery of super-mediators of information diffusion in social networks, Lecture Notes in Artificial Intelligence (Proc. of DS 2010), 査読有, Vol.6332, 2010, 144-158

K.Saito, M.Kimura, K.Ohara, and H.Motoda, Selecting information diffusion models over social networks for behavioral analysis, Lecture Notes in Artificial Intelligence (Proc. of ECML PKDD 2010), 査読有, Vol.6323, 2010, 180-195

K.Saito, M.Kimura, K.Ohara, and H.Motoda, Behavioral analyses of information diffusion models by observed data of social networks, Lecture Notes in Computer Science (Proc. of SBP 2010), 査読有, Vol.6007, 2010, 149-158

T.Fushimi, K.Saito, M.Kimura, H.Motoda, and K.Ohara, Finding relation between PageRank and voter model, Lecture Notes in Artificial Intelligence (Proc. of PKAW 2010), 査読有, Vol.6232, 2010, 208-222

Y.Yoshikawa, K.Saito, H.Motoda, K.Ohara, and M.Kimura, Acquiring expected influence curve from single diffusion sequence, Lecture Notes in Artificial Intelligence (Proc. of PKAW 2010), 査読有, Vol.6232, 2010, 273-287

K.Saito, M.Kimura, K.Ohara, and H.Motoda, Efficient estimation of cumulative influence for multiple activation information diffusion model with continuous time delay, Lecture Notes in Artificial Intelligence (Proc. of PRICAI 2010), 査読有, Vol.6230, 2010, 244-255

M.Kimura, K.Saito, R.Nakano, and H.Motoda, Extracting influential nodes on a social network for information diffusion, Data Mining and Knowledge Discovery, 査読有, Vol.20 (1), 2010, 70-97

木村昌弘、斉藤和巳、中野良平、元田浩、社会ネットワークにおける有力ノード抽出のための情報拡散モデルの学習、人工知能学会論文誌、査読有、Vol.25(1)、2010、215-223

M.Kimura, K.Saito, and H.Motoda, Blocking links to minimize contamination spread in a social network, ACM Transactions on Knowledge Discovery from Data, 査読有, Vol.3 (2), 2009, 9:1-9:23

M.Kimura, K.Saito, and H.Motoda, Efficient estimation of influence functions for SIS model on social networks, Proceedings of the Twenty-First International Joint Conference on Artificial Intelligence (IJCAI-09), 査読有, 2009, 2046-2051

K.Saito, M.Kimura, K.Ohara, and H.Motoda, Learning continuous-time information diffusion model for social behavioral data analysis, Lecture Notes in Artificial

- Intelligence (Proc. of ACML 2009), 査読有, Vol.5828, 2009, 322-337
K.Saito, M.Kimura, and H.Motoda, Discovering influential nodes for SIS models in social networks, Lecture Notes in Artificial Intelligence (Proc. of DS 2009), 査読有, Vol.5808, 2009, 302-316
M.Kimura, K.Saito, R.Nakano, and H.Motoda, Finding influential nodes in a social network from information diffusion data, Proceedings of the Second Workshop on Social Computing, Behavioral Modeling and Prediction (SBP 2009), 査読有, 2009, 139-145
T.Fushimi, T.Kawazoe, K.Saito, M.Kimura, and H.Motoda, What does an information diffusion model tell about social network structure?, Lecture Notes in Artificial Intelligence (Proc. of PKAW 2008), 査読有, Vol.5476, 2009, 122-136
M.Kimura, K.Saito, and H.Motoda, Solving the contamination minimization problem on networks for the linear threshold model, Lecture Notes in Artificial Intelligence (Proc. of PRICAI 2008), 査読有, Vol.5351, 2009, 977-984
- 21 K.Saito, M.Kimura, and H.Motoda, Effective visualization of information diffusion process over complex networks, Lecture Notes in Artificial Intelligence (Proc. of ECML PKDD 2008), 査読有, Vol.5212, 2008, 326-341
- 22 木村昌弘, 斉藤和巳, PMM 型主成分分析を用いた文書ストリームの主要潜在トピック抽出、日本応用数理学会論文誌、査読有、Vol.18(3)、2008、363-388
- 23 K.Saito, M.Kimura, and R.Nakano, Prediction of information diffusion probabilities for independent cascade model, Lecture Notes in Artificial Intelligence (Proc. of KES 2008), 査読有, Vol.5179, 2008, 67-75
- 24 M.Kimura, K.Yamakawa, K.Saito, and H.Motoda, Community analysis of influential nodes for information diffusion on a social network, Proceedings of the 2008 International Joint Conference on Neural Networks (IJCNN 2008), 査読有, 2008, 1359-1364
- 25 M.Kimura, K.Saito, and H.Motoda, Minimizing the spread of contamination by blocking links in a

network, Proceedings of the Twenty-Third AAAI Conference on Artificial Intelligence (AAAI-08), 査読有, 2008, 1175-1180

〔学会発表〕(計 44 件)

斉藤和巳、木村昌弘、大原剛三、元田浩、Behavioral analyses of information diffusion models by observed data of social network、日本ソフトウェア科学会ネットワークが創発する知能研究会JWEIN10ワークショップ、2010年8月20日、東京都、キャンパスイノベーションセンター東京

木村昌弘、斉藤和巳、中野良平、元田浩、情報拡散データに基づいた社会ネットワークのノードランキング、人工知能学会第23回全国大会、2009年6月18日、香川県、サンポートホール高松

木村昌弘、斉藤和巳、元田浩、社会ネットワークにおける汚染拡散最小化のためのリンク封鎖、人工知能学会第9回データマイニングと統計数理研究会、2009年3月4日、京都府、メルパルク京都

伊藤政志、木村昌弘、斉藤和巳、ブログ空間における情報伝搬パターンの分析、第7回情報科学技術フォーラムFIT2008、2008年9月3日、神奈川県、慶応大学

6 . 研究組織

(1)研究代表者

木村 昌弘 (KIMURA MASAHIRO)
 龍谷大学・理工学部・教授
 研究者番号：10396153

(2)研究分担者

中野 良平 (NAKANO RYOHEI)
 中部大学・工学部・教授
 研究者番号：90324467
斉藤 和巳 (SAITO KAZUMI)
 静岡県立大学・経営情報学部・教授
 研究者番号：80379544

(3)連携研究者

なし