

科学研究費助成事業（科学研究費補助金）研究成果報告書

平成 25 年 4 月 26 日現在

機関番号：14401

研究種目：若手研究（B）

研究期間：2009～2011

課題番号：21700166

研究課題名（和文） インターネット空間における影響伝播

研究課題名（英文） Influence Diffusion on the Internet

研究代表者

松村 真宏（MATSUMURA NAOHIRO）

大阪大学・大学院経済学研究科・准教授

研究者番号：10379159

研究成果の概要（和文）：インターネット空間における影響伝播のプロセスを表す数理モデル（IDM）を提案し、実際に様々なデータを分析することで提案手法の妥当性の検討を行った。また、IDM の理論解析も行い、データのゆらぎに頑健な指標を提案した。また、ビッグデータに対応できるように、簡便化した方式による IDM の近似解の導出を行った。

研究成果の概要（英文）：I proposed a mathematical model named IDM that represents an influence diffusion process on the Internet and investigated the validity of the model by analyzing the various data. Based on the theoretical analysis of the IDM, I also improved the IDM to be more reliable for the biased data. Also, I proposed a computationally less intensive IDM to be able to apply to big data.

交付決定額

（金額単位：円）

	直接経費	間接経費	合計
2009 年度	1,400,000	420,000	1,820,000
2010 年度	600,000	180,000	780,000
2011 年度	600,000	180,000	780,000
年度			
年度			
総計	2,600,000	780,000	3,380,000

研究分野：知能情報学

科研費の分科・細目：情報学・知能情報学

キーワード：影響伝播モデル、IDM

1. 研究開始当初の背景

インターネット空間はウェブページ/サイトがリンクで繋がったネットワーク状の構造を持っている。ここで、各ウェブページ/サイト上の話題がリンクを介して他のウェブページ/サイトに広まっていく様子を捉えることで、インターネット空間に広まる影響を定量的に測定することが可能となる。

申請者が取り組んでいる影響伝播モデルは話題の再帰的な伝播プロセスに着目して影響伝播を測定する手法である。従来の社会ネ

ットワーク分析ではネットワークの構造的特徴にしか注目していなかったため、表層的な分析しかできなかった。一方、影響伝播モデルを適用することにより、話題の伝播経路を推定できる、時系列データを扱える、話題の影響力を計測できる、話題間の関係を計測できる。といった新たな指標を計測できる。つまり、インターネット空間をウェブページ/サイトとリンクからなるネットワーク構造として捉えるのではなく、話題の伝播経路としてネットワーク構造を推定するため、話題

ごとに異なるネットワーク構造を扱うことが可能になる。これは、従来のネットワーク分析で対象としてきたウェブページとリンクからなるネットワーク構造とは根本的に異なるアプローチである。

このような特徴をもつ影響伝播モデルを大規模データに適用することにより、インターネット空間における影響伝播の時系列プロセスを、ネットワークの多面的な側面から明らかにすることに取り組む。

2. 研究の目的

世論や流行の形成において影響伝播の相転移を起こすための条件を、インターネット空間の大規模データを用いて実証的に解析的に求めることを試みる。ここで対象とする世論や流行の形成は、話題の伝播経路と伝播範囲に対応しており、これらは影響伝播モデルにより求まるインターネット空間のネットワークの多面的構造によって表される。このネットワークの多元構造に対して、影響伝播モデルによる話題の影響力、およびその伝播経路と伝播範囲、影響伝播モデルによる重み付き有向ネットワークの構造的特徴（クラスタリング係数、平均最短経路長、中心性、次数など）などの指標を用いることで、世論や流行の形成段階を特徴づける条件を解き明かすことに取り組む。

3. 研究の方法

影響伝播モデルの構築と理論的解析、近似モデルの構築を行った。また、モバイル SNS サイトにおける紹介データ・行動データ、Twitter の大規模データを収集した。また、提案した分析アルゴリズムを用いて実際にデータを収集し、考察を行った。

4. 研究成果

(1) 影響伝播モデルの線形代数表現による拡張を行った。具体的には、影響伝播モデルを再帰構造行列と重み行列のシュア積で表すことにより、影響伝播モデルの理論的背景の解明を行った。その結果、汎用的な統計解析ソフトウェアによって効率良く影響伝播モデルの計算が行えるようになった。以下に計算方法を簡単に示す。

まず、ネットワークの再帰的な伝播構造を表す再帰構造行列 M^* を以下のように求める (M は有向接続行列)。

$$M^* = M + \beta M^2 + \beta^2 M^3 + \dots + \beta^{n-1} M^n$$

また、重み行列 A^* を以下のように定義する (A はメッセージ×語からなる行列)。

$$A^* = AA^T$$

M^* と A^* を用いて影響量行列 P を以下のように

に定義する (演算子 \circ はシュア積を表す)。

$$P = M^* \circ A^*$$

この行列 P がメッセージ i からメッセージ j に伝播する影響量を現しており、行列計算のみで算出できる。また、メッセージ間の影響量だけでなく、投稿者間の影響料や語と語の間の影響量も算出できる。以下に、とある製品発表に関するツイートデータ (文書数 5890、リンク数 1890、語の種類数 649) を提案手法で分析した結果を図 1 に示す。

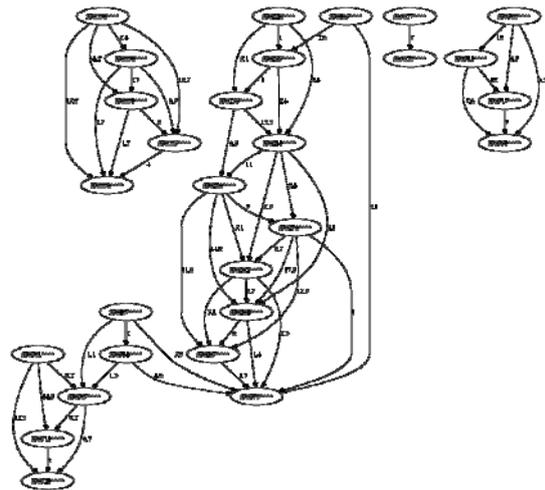


図 1. ツイート間の影響伝播ネットワーク図

(2) IDM ではネットワーク上を伝播した語だけが影響量として計上されるので、文脈に関係のない語の影響を受けにくいアルゴリズムになっている。しかし、語の頻出頻度が高くなれば偶然に伝播する可能性も高くなるため、高頻度語は本来の影響量より高く計上される傾向がある。そこで、ネットワーク構造 (ノード数とリンク数) と語の出現頻度から語の影響量の期待値を求めた。

途中の計算式は省略するが、最終的に得られる語の影響力の近似式は

$$E_w = f_w R_w R_L / (1 - \beta R_w R_L)$$

になる。ただし、メッセージ数を N 、リンク数を L 、語 w の文書頻度を f_w とすると、メッセージに語 w が出現する割合 R_w は $R_w = f_w / N$ となる。また、メッセージに接続されているリンクの割合 R_L は $R_L = L / (N - 1)$ となる。

出現頻度の高い語は影響量が大きくなる傾向があるので、影響量は出現頻度に応じた正規化が必要がある。そこで、語 w の影響量 k_w 、期待値を E_w とすると、影響量の I/E 比 (influence-to-expectation ratio) は以下のように定義できる。

$$I/E \text{ 比} = k_w / E_w$$

I/E 比は 1 より大きくなるほど文脈に乗って語が伝播していることを表しているので、影響量のより正確な見積もりになる。ツイートの影響量、期待値、I/E 比を算出した結果を

表 1 に示す。

Tweet ID	影響量	RE 数	語種	期待値	I/E
558090****	106	4	4	0.03	3905
558701****	67	3	5	0.02	3467
558276****	64	2	9	0.46	138
569677****	61	2	5	0.36	170
559511****	61	2	10	0.47	130

表 1. 影響量の上位 5 ツイート

(3) 計算量の見積も行い、大規模データにも適用できることを示した。導出過程は省略するが、最終的に $P = M * A *$ の計算量の見積もりは $O(kmn) + O(mk^n)$ になる。

(4) 感度分析

影響量の期待値は語の出現箇所とリンク構造がランダムの場合であったが、これらが偏っていれば語の影響量の期待値も異なる。そこで、語の出現箇所のランダムネスを表すパラメータ $\alpha_w (0 \leq \alpha_w \leq 1)$ 、リンク構造のランダムネスを表すパラメータ $\alpha_L (0 \leq \alpha_L \leq 1)$ を導入し、影響量の期待値を拡張する。 α_w 、 α_L が大きくなるほどランダムさが増すとすると、語の出現する可能性のあるノード数 N_w とリンクの貼られる可能性のあるノード数 N_L は以下のように表せる。

$$N_w = (N - f_w) * \alpha_w + f_w$$

$$N_L = (N - 1 - L) * \alpha_L + L$$

N_w 、 N_L を用いると、メッセージに語 w が出現する割合 R'_w は $R'_w = f_w / N_w$ 、メッセージに接続されているリンクの割合 R'_L は $R'_L = L / N_L$ となる。このとき、 $R' = R'_w / R'_L$ と置くことで、式(20)の影響量の期待値は以下のように拡張できる。

$$E_w = f_w R' / (1 - \beta R')$$

$\beta R' = 1$ のときは $E_w = f_w (f_w - 1) / 2$ になる。ここで、ランダムネスを変えたときの影響量の変化量を見るために、 $N=1000$ 、 $\beta=1$ 、 $0 \leq f_w \leq 500$ 、 $0 \leq L \leq 1000$ における影響量の期待値を図 2 に示す。図 2 中の 4 枚の曲面は、それぞれ下から順に $\alpha_w = \alpha_L = 1$ 、 $\alpha_w = \alpha_L = 0.8$ 、 $\alpha_w = \alpha_L = 0.6$ 、 $\alpha_w = \alpha_L = 0.4$ のときの影響量の期待値を表している。これより、語の出現箇所やリンク構造が偏るほど影響量も大きくなる事が分かる。また、その傾きは語の出現件数 f_w やリンク数 L が大きくなるほど急激に大きくなる。

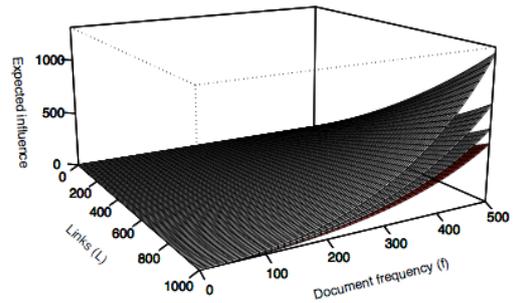


図 2. 語の出現箇所の偏りとリンク構造の偏りに対する影響量の期待値

(5) 近似モデル

ランダムネットワークを想定したときの IDM 影響力の期待値を以前求めたが、ランダムネットワークを実際のネットワーク構造を反映させたものに置き換えることで IDM 影響量を確率的に求めることが可能になる。具体的には、どの語がどの文書に登場するかが分かり、文書ごとにどれくらい拡散する力を持っているかが分かればよい。前者は転置インデックスから求め、後者は隣接行列から確率的に求まる。そこで、語 w を含む文書集合を D_w 、語 w の頻度を f_w 、全文書数を N 、隣接行列 M のノイマン級数における文書 d の出次数の和を l_d とすると、語 w の IDM 影響力 $I(w)$ は以下の式で求まる。この手法を IDMstat と呼ぶ。

$$I_w = f_w (f_w - 1) / (N(N-1)) \sum_{d \in D_w} \sum_i i^n l_d$$

また、同じノード数 N 、リンク数 L 、語の出現頻度 f_w における語 w の IDM 影響力の期待値 E_w (上述) を使えば、 I_w / E_w の大きな語 w が特徴的な語ということになる。

ツイートデータを使って 100 回計算したときの処理時間を測ったところ、IDM (従来の行列計算による方法) では 0.317 秒かかったのに対して、IDMstat では 0.071 秒であり、4.5 倍程度速くなったことが確認できた。なお、IDMstat はデータ数に対して線形にしか計算量が増えず、保持するデータ量も少なく済むので、データ数が増えればもっと顕著な差が出てくる。

また、行列計算で得られる影響量の厳密解を gold standard としたときの IDMstat の精度評価を行った。IDM と文書頻度 DF と IDMstat の結果の順位相関係数の結果を図 3 に示すが、期待通り IDMstat は DF よりも IDM に近い結果になることが確認できた。

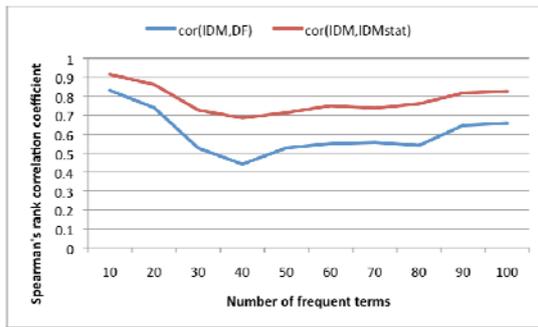


図 3. IDM と DF と IDMstat の結果の順位相関係数

(6) モバイル SNS を対象としてユーザの金銭的価値の研究を行った。従来、このようなサイトにおけるユーザの価値は、ユーザの消費行動（具体的には、来訪問隔、来訪頻度、金額の 3 指標）、もしくはユーザの友人関係のネットワーク分析によってなされてきた。本研究では、これらの指標を統合し、ユーザの友人関係および友人たちの消費行動を同時に考慮した新しい価値指標を提案した。モバイル SNS 運営会社から提供されたデータに適用し、提案手法の評価を行った。

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 1 件)

- ① 松村真宏、山本品：ブログ空間におけるインフルエンサーおよび消費者インサイトの発見、季刊マーケティングジャーナル、30(3)、pp. 82-94 (2010)

[学会発表] (計 16 件)

- ① Hikaru Yamamoto and Naohiro Matsumura: Optimal Heterophily for Word-of-Mouth Diffusion, International AAAI Conference on Weblogs and Social Media (ICWSM-09)
- ② Hikaru Yamamoto and Naohiro Matsumura: The Power of Grassroots Influentials, The 2009 IEEE International Conference on Social Computing (SocialCom 2009)
- ③ Naohiro Matsumura and Hikaru Yamamoto: Monetary Value of Customer Networks in Mobile Social Networking Services, International AAAI Conference on Weblogs and Social Media (ICWSM2010)
- ④ Hikaru Yamamoto and Naohiro Matsumura: Measuring the Value of Customer in Mobile Social Networking Services, The

2010 INFORMS Marketing Science Conference

- ⑤ Hikaru Yamamoto, Naohiro Matsumura: Marketing Ecosystem: The Dynamics of Twitter, TV Advertising, and Customer Acquisition, Social Mobile Web (SMW'11)
- ⑥ Hikaru Yamamoto, Naohiro Matsumura: The Effect of Relational Context on Personal Influence, Workshop on Social Behavioral Analysis and Behavioral Change (SBABC2011), The Third IEEE International Conference on Social Computing (SocialCom2011)

[図書] (計 2 件)

- ① 松村真宏、三浦麻子：人文・社会科学のためのテキストマイニング、誠信書房 (2009)
- ② 西田豊明・角康之・松村真宏：社会知デザイン、オーム社 (2009)

[その他]

ホームページ等
<http://mtmr.jp>

6. 研究組織

(1) 研究代表者

松村 真宏 (MATSUMURA NAOHIRO)
 大阪大学・大学院経済学研究科・准教授
 研究者番号：10379159