

時間遅れパラメータ付き SIS モデルの期待影響度分析*

小出 明弘† 斉藤 和巳† 元田 浩‡ 木村 昌弘§ 大原 剛三¶

静岡県立大学† 大阪大学‡ 龍谷大学§ 青山学院大学¶

1. はじめに

近年, ソーシャルネットワークを通じた情報拡散現象に対する関心が急速に高まっている. 一般に, ソーシャルネットワーク上の情報拡散過程は, ノードが非アクティブ(情報が伝わっていない状態)からアクティブ(情報が伝わった)に変化する過程としてモデル化され, IC モデル・LT モデルを用いた影響最大化問題や汚染最小化問題等の研究がおこなわれている[1, 2].

私たちは IC, LT モデルに用いられる SIR モデルを拡張した SIS モデルについての研究を行っている[3]. SIS モデルでは, S 状態にあるブログユーザーが, 記事を読んで触発され, 記事を投稿するまでの I 状態に移行し, その後同じトピックの記事を読む S 状態に戻り, また投稿するという過程が繰り返される. こうした SIS モデルにおける情報拡散過程は, SIR のものと大きく異なるものである. さらに現実の情報拡散では, S 状態にあるブログユーザーが記事に触発されて投稿するまでの間には時間遅れが発生するものと考えられる. そこで私たちは, SIS モデルをより現実に近付けた連続時間遅れ付き SIS モデルを提案する. また, SIS モデルを用いた期待影響度分析では, 従来シミュレーション回数を設定し, その回数分のシミュレーションを行うことで期待影響度を出力していた. よって, 従来法ではシミュレーション回数分の処理が必要であり, 計算負荷が問題となっている. 本論文では, SIS モデルを用いた期待影響度分析において新計算法を提示し, 実際にどれ位の処理時間の差となるか分析する.

2. 分析方法

2.1 新計算法

連続時間遅れ付き SIS 型 IC モデルを対象とし, 各ノードの期待影響度を求める新計算法の概要について説明する. 簡略化のため, IC モデルの拡散確率 p と, 情報拡散の時間遅れを規定する指数分布のパラメータ r はすべてのリンクで一定とする. また, 与えられた無向ネットワークの全ノード集合を $V = \{1, \dots, N\}$ とし, その隣接行列を A とする. 隣接行

列の第 (i, j) 成分の $a_{i,j}$ は, ノード i と j 間にエッジがあれば 1 に, なければ 0 に設定する.

各ノードを情報源とするとき, J ステップの情報拡散が成功することでアクティブとなるノード数の期待値ベクトルは $(pA)^J \mathbf{1}$ である. ここで, $\mathbf{1}$ は全要素が 1 の N 次元ベクトルを表す. いま, 時刻 0 から T までにアクティブとなった積算ノード数の期待値で影響度を定義する. このとき, J ステップの情報拡散が時間間隔 $[0, T]$ の範囲で起こる確率は以下となる.

$$R(J; T) = 1 - \exp(-rT) \sum_{j=1}^J \frac{(rT)^{j-1}}{(j-1)!}$$

よって, 時間間隔 $[0, T]$ での各ノードの積算影響度の期待値ベクトルは次式で求めることができる.

$$\sigma(T) = \sum_{j=1}^{\infty} R(J; T) (pA)^j$$

実際には, $R(J; T) (pA)^j > \varepsilon$ となる範囲での和を求め, 各ノードの積算影響度の期待値ベクトルとする.

2.3 分析設定

本分析では, 従来法による期待影響度分析と新提案法による期待影響度分析の処理速度をプロットし, 2 つの手法の処理速度を比較する. 従来法では, モデルの定義に従い, シミュレーションで期待影響度を求める. このときのシミュレーション回数を 1, 10, 100, 1000, 10000 に設定する. また, モデルの指数パラメータは $r = 1$ とし, 期待影響度を求める最終時刻は $T = 10$ に設定する. 一方, 新計算法で期待影響度を計算する範囲を規定するパラメータは, 十分に小さく, $\varepsilon = 10^{-12}$ に設定する.

3. 実験

3.1 実験データ

本研究では 2 つのネットワークデータを用いて分析する.

1 つ目は, ブログのトラックバックネットワーク (以下ブログデータと呼ぶ) のデータで, ノード数は, 12,047 でリンク数は, 79,920 である.

2 つ目は, 日本の「ウィキペディア」内の「人名一覧」からの人物ネットワーク (以下ウィキデータと呼ぶ) で, ノード数は, 9481 でリンク数は, 245,044 である. これらのネットワークには, 次数分布がべき則分布に従うなど多くの大規模ネットワークと同様な特徴が見られる.

*Analysis of Influence Calculation on SIS Model with Continuous-time Delay

†Akihiro Koide, Kazumi Saito · University of Shizuoka

‡Hiroshi Motoda · Osaka University

§Masahiro Kimura · Ryukoku University

¶Kozou Ohara · Aoyama Gakuin University

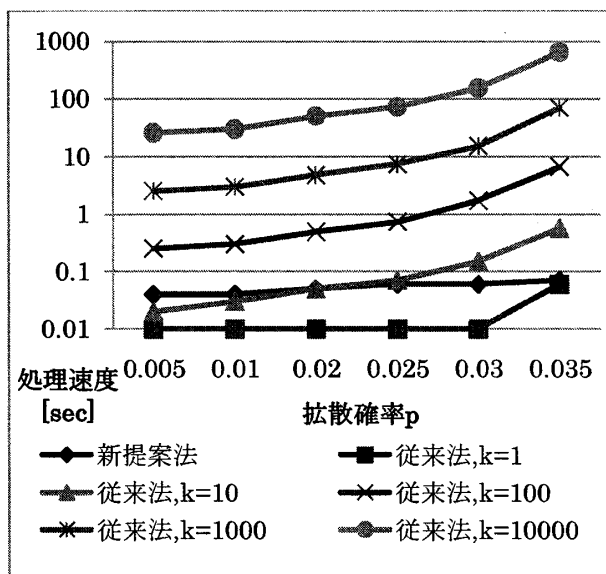


図 1 : ブログデータによる分析結果

ブログデータにおいては、拡散確率 $p = 0.005, 0.01, 0.02, 0.025, 0.03$ および 0.035 に設定する。なお、ブログデータにおける最大固有値の逆数はおよそ 0.025 である。一方、ウィキデータにおいては拡散確率 $p = 0.005, 0.006, 0.0067, 0.007$ および 0.008 とする。なお、ウィキデータにおける最大固有値の逆数はおよそ 0.0067 である。

3.2 分析結果・考察

分析結果を図1, 2に示す。ここで k は試行回数を表している。従来法では、シミュレーション回数100回までは処理速度がそれほど大きくなることはないが、1000回を超えると p が大きくなるにつれ計算負荷が大きくなっていることが分かる。正確な期待影響度を求めるには、それだけ多くのシミュレーション回数が必要であると考えられるため、現実にシミュレーションを行うにはかなりの計算負荷がかかる。また、ブログデータでは $p = 0.035$ 、ウィキデータでは $p = 0.008$ において計算負荷が急激に大きくなっていることが分かる。これは、拡散確率 p が最大固有値の逆数を超えたことで起こったことであることが示唆される。

一方、新提案法では従来法に比べ処理速度がかなりはやいことがわかる。また、拡散確率を大きくしても計算負荷はほとんど変化しない。また、従来法では拡散確率における期待影響度を正確に求めることができないのに対して、新提案法では求めることができるようになってきている。これらの結果から、従来法と比べ新提案法による期待影響度分析のほうがより正確で計算負荷の小さい手法であることが示唆される。

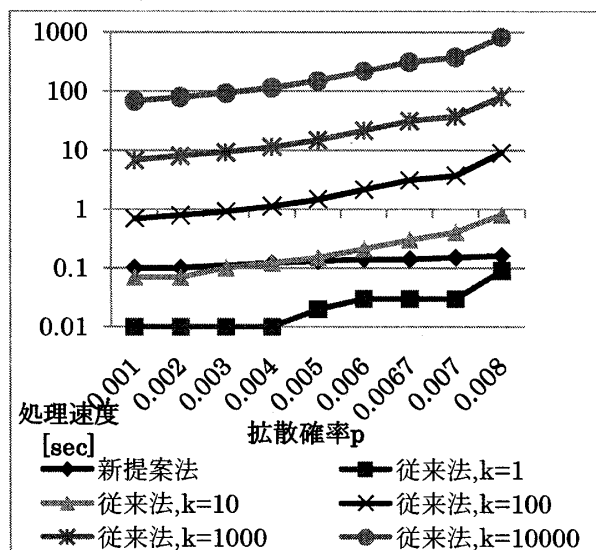


図 2 : ウィキペディアデータによる分析結果

4. おわりに

本論文では、従来の期待影響度分析法とは異なる新提案法を提示し、互いの計算負荷比較を行った。そして、分析結果から新提案法は正確さ、並びに計算負荷の両方の面において従来法よりも優れているという知見を得られた。

今後は、この新提案法を用いて影響最大化問題・汚染最小化問題についての研究を行っていきたい。

参考文献

- [1] Kempe, D, Kleinberg, J. and Tardos, E., "Maximizing the spread of influence through a social network", Proc. KDD'03, pp. 137-146, (2003).
- [2] Kimura, M., Saito, K. and Motoda, H., "Blocking Links to Minimize Contamination Spread in a Social Network," ACM Transactions on Knowledge Discovery from Data, Vol.3, No.2, Article 9, pp.1-23, 2009.
- [3] 小出 明弘, 水本 嗣留, 斉藤 和巳, 元田 浩, 木村 昌弘, "SIS モデルの影響度とコミュニティ構造の関係分析," 情報科学技術フォーラム (FIT) 2009.