1.14-1

価値重みつき混合投票者モデルを用いた 反多数派オピニオニストの抽出

Detecting Anti-majority Opinionists Using Value-weighted Mixture Voter Model

木村昌弘*1 斉藤和巳*2 大原剛三*3 元田浩*4 Masahiro Kimura Kazumi Saito Kouzou Ohara Hiroshi Motoda

*3青山学院大学 *2静岡県立大学 *4大阪大学 *1龍谷大学 Ryukoku University University of Shizuoka Aoyama Gakuin University Osaka University

We address the problem of detecting anti-majority opinionists using the value-weighted mixture voter (VwMV) model. This problem is motivated by the fact that some people do not always agree with the majority and support

the minority. We extend the value-weighted voter model to include this phenomenon with the anti-majoritarian tendency of each node as a new parameter, learn this parameter as well as the value of each opinion from a sequence of observed opinion data over a social network. We experimentally show that it is possible to learn the anti-majoritarian tendency of each node correctly as well as the opinion values, whereas a naive approach which is based on a simple counting heuristic fails. We also show theoretically that, in a situation where the local opinion share can be approximated by the average opinion share, it is not necessarily the case that the opinion with the highest value prevails and wins when the opinion values are non-uniform, whereas the opinion share prediction problem becomes ill-defined and any opinion can win when the opinion values are uniform. The simulation results

1. はじめに

ソーシャルメディアの登場により、Web 空間内に様々な大 規模ソーシャルネットワークが形成されてきている。それら ソーシャルネットワークは、ニュース、アイデア、オピニオン、 噂など、多様な情報の拡散を媒介するので、我々の生活にも大 きな影響を及ぼしうる [Newman 02, Domingos 05, Wu 08]。 したがって、近年、ソーシャルネットワークにおける情報拡散 の研究が注目されている [Gruhl 04, Leskovec 07a]。中でも、 与えられたソーシャルネットワークにおいて、情報を普及させ る際に有効な、影響力のある限られた数のノード群を見つける という、"影響最大化問題"は非常によく研究されており、そ の近似解を求める有効な多くの新しいアルゴリズムが提案さ れている [Kempe 03, Leskovec 07b, Chen 09, Kimura 10a]。 しかしながら、これらの研究で用いられたモデルでは、影響の 広がりに焦点をあてているので、ノードはアクティブか非アク ティブという2状態の一方のみをとることになっていた。

support that this holds for typical real world social networks.

ユーザが複数の選択肢の中から一つを選ぶことができる競 合的オンラインサービスのような応用では、モデルはノード の複数の状態を扱えなければならないので、異なるアプロー チが要求される。また、選択に強い影響を与えるものとして、 Even-Dar 07, Yang 09]。投票者モデルは、ノードがその近傍 のオピニオンに基づいてそのオピニオンを変更する、最も基本 的な確率過程モデルである。我々は、文献 [Kimura 10b] にお いて、オピニオン価値を含むように投票者モデルを拡張し、過 去のオピニオン拡散の観測データから将来のオピニオンシェア を予測する問題を論じた。また、理論的解析とシミュレーショ ン実験により、オピニオンの期待シェア予測問題においては、

品質、ブランド、および権威など、各選択の価値を考慮に入れ ることも重要である。上記応用はオピニオンの拡散・形成と同 じ問題クラスに含まれるが、このような問題の分析に最もよく 使われるモデルの一つとして、投票者モデルがある [Sood 05, オピニオン価値が非一様であるとき、十分が時間が経てばオピ ニオン価値が最大のオピニオンが一人勝ちする(コンセンサス が得られる)ことを示した。

本論文では、異なった角度からオピニオン拡散・形成の問題 に取り組む。まず、通常の投票者モデルでは、人々は近傍の多 数派オピニオンに従う傾向があると仮定されていることに注意 する。しかしながら現実世界では、多数派には同意せず少数派 のオピニオンを支持する人が常に存在する。我々は、このよう な人々の存在がオピニオンシェアにどのような影響を及ぼすの かに興味をもち、"複数オピニオンをもつ価値重みつき投票者 モデル" [Kimura 10b] を、各ノードの反多数派傾向を新たな パラメータとして組み込みことで反多数派の効果を含むよう に拡張する。反多数派の概念を導入したのは我々が最初ではな く、"反投票者モデル"と呼ばれる2つのオピニオンをもつオ ピニオン拡散モデルがすでに存在し、その統計的性質が研究さ れている [Röllin 07]。ここに、反投票者モデルとは、各人がラ ンダムにその近傍の一人を選び、その人と反対のオピニオンを とるというオピニオン拡散モデルである。我々は、単純な反投 票者モデルを複数オピニオンをもつ価値重みつき反投票者モデ ルに拡張し、それを複数オピニオンをもつ価値重みつき投票者 モデルと線形結合させる。構築したモデルは、元のパラメータ (オピニオン価値)に加えて、ノードごとに新たなパラメータ (反多数派傾向の測度である価値重みつき反投票者モデルに対 する重み)をもっている。このオピニオン拡散モデルを"価値 重みつき混合投票者 (VwMV) モデル"と呼ぶ。

我々は、反多数派傾向とオピニオン価値の両方のパラメータ を、オピニオン拡散の観測データから学習する手法を提案し、 実ソーシャルネットワークトポロジーを用いた実験でその有効 性を実証する。また、VwMV モデルでは、オピニオン価値が 非一様でも、オピニオン価値が最大のオピニオンが勝者になる とは限らないということを、近傍での期待局所オピニオンシェ アがネットワーク全体での期待オピニオンシェアによって近似 できる場合に理論的に示す。

連絡先: 木村昌弘, 龍谷大学理工学部電子情報学科, 〒520-2194 大津市瀬田大江町横谷 1-5, kimura@rins.ryukoku.ac.jp

2. 価値重みつき混合投票者モデル

価値重みつき混合投票者 (VwMV) モデルを定義する。G=(V, E) を自己リンクをもつ無向(双方向)ネットワークとす る。ここに、V はノード全体の集合で E $(\subset V \times V)$ はリン ク全体の集合である。 ノード $v \in V$ の近傍ノード全体の集合 を、 $\Gamma(v) = \{u \in V; (u,v) \in E\}$ とする。 $v \in \Gamma(v)$ に注意。 整数 K(>2) に対して、G 上での K 個のオピニオン (オピ ニオン $1, \dots,$ オピニオン K) の広がりを考える。ここに、各 ノードは、時刻 t=0 で K 個のオピニオンの一つを等確率 で持ち、任意の時刻 t(>0) では一つのオピニオンを持つとす る。 $f_t:V \to \{1,\cdots,K\}$ を時刻 t での "オピニオン分布" と する。ここに、 $f_t(v)$ は時刻 t でのノード v のオピニオンを表 す。 fo は初期オピニオン分布を表していることに注意。時刻 t にノード $v \in V$ のオピニオン更新が行われるとき、任意の $k \in \{1, 2, \dots, K\}$ に対して、時刻 t の直前にオピニオン k を 持っている、ノードvの近傍ノード数を $n_k(t,v)$ とする。すな わち、

$$n_k(t, v) = |\{u \in \Gamma(v); \ \varphi_t(u) = k\}|$$

である。ここに、 $\varphi_t(u)$ はノード u が時刻 t の直前にもっているオピニオンを表す。

 $V_{
m WMV}$ モデルにおけるオピニオン拡散過程は次のとおりである。

- 1. 時刻 0 で各ノード v は、パラメータ 1 の指数分布 *1 のようなある確率分布にしたがって、独立にそのオピニオン更新時刻 t を決める。引き続き、次のオピニオン更新時刻は、各更新時刻 t において同様に決定される。
- 2. 更新時刻 t において、ノード v は、次の確率分布にしたがって K 個のオピニオンの中から一つのオピニオンを選択する。

$$P(f_t(v) = k) = (1 - \alpha_v) p_k(t, v, \boldsymbol{w})$$

+ $\alpha_v \frac{1 - p_k(t, v, \boldsymbol{w})}{K - 1}, \quad (k = 1, \dots, K) \quad (1)$

ただし、

$$p_k(t, v, \mathbf{w}) = \frac{w_k n_k(t, v)}{\sum_{j=1}^K w_j n_j(t, v)}, \quad (k = 1, \dots, K) \quad (2)$$

であり、 $\mathbf{w}=(w_1,\cdots,w_K)$ である。ここに、 $w_k (>0)$ はオピニオン k の "オピニオン価値" を表すパラメータであり、 $\alpha_v (0 \leq \alpha_v \leq 1)$ はノード v の "反多数派傾向" を表すパラメータである。

3. 本過程は、初期時刻 t=0 から、次の更新時刻が与えられた最終時刻 T を過ぎるまで繰り返される。

ここに、 $V_{\rm W}MV$ モデルにおいては、式 (1), (2) より、ノード v は、 α_v の値が小さいならば多数派オピニオニストのように振る舞い、 α_v の値が大きいならば反多数派オピニオニストのように振る舞うことに注意する。

3. 学習問題と挙動解析

ネットワーク G 上の $V{
m wMV}$ モデルを期間 [0,T] における 観測データ \mathcal{D}_T から同定する問題を考える。ここに、 \mathcal{D}_T は、

*1 平均遅れ時間は1となる。

 $0 \le t \le T$ なる時刻 t にノード v がそのオピニオンをオピニオン k に更新したことを表す (v,t,k) の全体である。同定されたモデルは、将来において各オピニオンのシェアがどのようになるかを予測するために、また、反多数派傾向の強いノード (反多数派オピニオニスト) や反多数派傾向の弱いノード (多数派オピニオニスト) を検出するために、利用可能であると考えられる。以下に、VwMV モデルの基本的性質を調べ、そのパラメータであるオピニオン価値 w_k $(k=1,\cdots,K)$ と反多数派傾向 α_v $(v \in V)$ の値を高精度に推定することは、オピニオンシェアを予測する上で重要であることを実証する。

任意のオピニオン k に対して、時刻 t でのその期待 "シェア" を $g_k(t)$ とする。すなわち、

$$g_k(t) = \left\langle |\{v \in V; f_t(v) = k\}| \middle/ \sum_{j=1}^K |\{v \in V; f_t(v) = j\}| \right\rangle$$

とする。十分時間 t が経った後での期待オピニオンシェア $g_k(t)$ の挙動を調べよう。統計物理学分野での研究 [Sood 05] にしたがって平均場アプローチ、

$$\frac{dg_k(t)}{dt} = (1 - g_k(t)) P_k(t) - g_k(t) (1 - P_k(t)), \quad (3)$$

 $(k=1,\cdots,K)$ を考える。ただし、 $P_k(t)$ は時刻 t でノードがオピニオン k を採用する確率である。ここで、ノード v の近傍での期待局所的オピニオンシェア $\langle n_k(t,v)/\sum_{j=1}^K n_j(t,v)\rangle$ は、期待オピニオンシェア $g_k(t)$ で近似できると仮定する。このとき、式 (1) から、

$$P_k(t) = (1 - \alpha)\,\tilde{p}_k(t, \boldsymbol{w}) + \alpha\,\frac{1 - \tilde{p}_k(t, \boldsymbol{w})}{K - 1},\tag{4}$$

 $(k=1,\cdots,K)$ という近似を得る。ここに、

$$\tilde{p}_k(t, \boldsymbol{w}) = \frac{w_k g_k(t)}{\sum_{j=1}^K w_j g_j(t)},$$
(5)

であり、 α は反多数派傾向 α_v $(v \in V)$ の平均値である。G が完全ネットワークであり $\alpha_v = \alpha$ $(\forall v \in V)$ ならば、式 (4), (5) は厳密に成立することに注意。

まず、オピニオン価値が一様である、すなわち、 $w_1=\cdots=w_K$ である場合を考える。このとき、式 (3),(4),(5) より、任意のオピニオン k に対して、 $\lim_{t\to\infty}g_k(t)=1/K$ が成り立つことが示される。

次に、オピニオン価値が非一様である場合を考える。オピニオン価値が最大のオピニオンを k^* とする。簡単のため、 $k \neq k^*$ ならば $w_k = w'$ $(< w_{k^*})$ と仮定する。ここに、w' は正定数である。 $g_1(0) = \cdots = g_K(0) = 1/K$ とする。このとき、式 (3), (4), (5) より、次が成り立つことが示される。

- 1. $\alpha < 1-1/K$ ならば、 $g_{k^*}(t) > 1/K$ (t>0) である、すなわち、オピニオン k^* は多数派になることが期待される。
- 2. $\alpha=1-1/K$ ならば、任意のオピニオン k に対して、 $g_k(t)=1/K$ $(t\geq 0)$ である。すなわち、任意のオピニオンが多数派になりうる。
- $3. \ \alpha > 1 1/K$ ならば、 $g_{k^*}(t) < 1/K \ (t>0)$ である。 すなわち、オピニオン k^* は少数派になることが期待される。

我々は上記の結果を、いくつかの実ソーシャルネットワークトポロジーを用いたシミュレーション実験で検証した。ページ数制限のため具体的な結果を載せることはできないが、すべての場合に理論どおりの挙動を示していたことに注意する。

4. 学習法

オピニオン拡散の観測データ \mathcal{D}_T から VwMV モデルのパラメータである、オピニオン価値 w_k $(k=1,\cdots,K)$ と反多数派傾向 α_v $(v\in V)$ の値を推定する手法を与える。本モデルのオピニオン拡散過程 (式(1)) に基づいて、対数尤度関数を、

$$\mathcal{L}(\mathcal{D}_T; \boldsymbol{w}, \boldsymbol{\alpha}) = \log \left(\prod_{(v,t,k) \in \mathcal{D}_T} P(f_t(v) = k) \right)$$
(6)

と定義できる。ここに、 $m{w}=(w_1,\cdots,w_K)$ であり、 α は $(\alpha_v)_{v\in V}$ なる |V| 次元ベクトルである。このとき、我々のパラメータ推定問題は、目的関数 $\mathcal{L}(\mathcal{D}_T;m{w},\alpha)$ を $m{w}$ と α に関して最大化する問題として定式化される。

まず、 α に依存するパラメータ、

$$\beta_{v,k,j}(\boldsymbol{\alpha}) = \left\{ \begin{array}{ll} 1 - \alpha_v & \text{if} \quad j = k, \\ \alpha_v / (K - 1) & \text{if} \quad j \neq k, \end{array} \right.$$

 $(v\in V,\;k,j=1,\cdots,K)$ を導入する。このとき、式 (2) より $1-p_k(t,v,{m w})=\sum_{j
eq k}p_j(t,v,{m w})$ であるので、 $P(f_t(v)=k)$ の定義 $(式\ (1))$ から、式 (6) は、

$$\mathcal{L}(\mathcal{D}_T; \boldsymbol{w}, \boldsymbol{\alpha}) = \sum_{(v, t, k) \in \mathcal{D}_T} \log \left(\sum_{j=1}^K \beta_{v, k, j}(\boldsymbol{\alpha}) \, p_j(t, v, \boldsymbol{w}) \right)$$

と表現できる。したがって、我々は、 $\rm EM$ アルゴリズム的な手法で w と α の値を推定する。本論文では、ページ数制限のため、アルゴリズムの詳細は省略する。

5. 実験評価

我々は、いくつかの大規模な実ソーシャルネットワークトポロジーを用いた実験により、提案学習法の有効性を実証した。本論文では特に、Wikipedia ネットワークデータにおける、ノードの反多数派傾向の推定誤差の結果と、反多数派傾向の高いノード群(反多数派オピニオニスト達)の抽出精度に関する結果を示す。ここに、本ネットワークデータは、ソーシャルネットワークのトポロジーがもつ顕著な特徴を備えたネットワークとして、文献 [Kimura 09] で利用されているものであり、ノード数 9,481、リンク数 245,044 の双方向ネットワークである。

5.1 実験設定

まず、各ノード $v\in V$ の真の反多数派傾向 α_v^* を、パラメータ a=2,b=2 のベータ分布からサンプルした。ここに、 α_v^* の平均値は 0.5 になること $(\alpha=0.5)$ が期待されていることに注意。次に、真のオピニオン価値を、 $w_k^*=5-4(k-1)/(K-1),(k=1,\cdots,K)$ と設定した。すなわち、 $w_1^*=5,w_2^*=5-4/(K-1),\cdots,w_K^*=1$ である。さらに、与えられたネットワークにおいて、各ノードの初期オピニオンを一様ランダムに選択し、真の VwMV モデルに基づいて期間 [0,T] のオピニオン拡散データ \mathcal{D}_T を生成した。そして、観測データ \mathcal{D}_T から各ノードの反多数派傾向を推定する問題を考えた。ノードの反多数派傾向の推

定誤差を推定誤差 $\mathcal E$ で、反多数派傾向の高いノード群の抽出精度を $\mathcal F$ 値 $\mathcal F$ で、それぞれ測定した。ここに、 $\mathcal E$ と $\mathcal F$ は、

$$\mathcal{E} = \frac{1}{|V|} \sum_{v \in V} |\hat{\alpha}_v - \alpha_v^*|, \quad \mathcal{F} = \frac{2|\hat{A} \cap A^*|}{|\hat{A}| + |A^*|}$$

で定義される。ただし、各 $\hat{\alpha}_v$ は推定したノード v の反多数派傾向であり、 A^* は真の反多数派傾向の高い上位 15% のノード群、 \hat{A} は推定された反多数派傾向の高い上位 15% のノード群である。

5.2 比較法

ノードの反多数派傾向の推定において、VwMV モデルにオピニオン価値を組み込むことの意義を調べるために、すべてのオピニオン価値が同じ値をとり、各ノードの反多数派傾向のみがパラメータとなるような、制限 VwMV モデルを比較法とした。本手法を、"一様価値 (uniform) 法"と呼ぶことにする。また、ノードの反多数派傾向を、そのノードの選択したオピニオンが近傍における少数派オピニオンであった回数を単純に数えることにより推定するという手法も比較法とした。本手法を、"ナイーブ (naive) 法"と呼ぶことにする。

5.3 実験結果

K の値が小さい場合 (K=3) と K の値が大きい場合 (K=10) を調べた。図 1 は、各手法の推定誤差 $\mathcal E$ を学習期間 T の関数として示している。図 2 は、各手法の F 値 $\mathcal F$ を観測期間 T の関数として示している。ここに、同じ実験を独立に 5 回繰り返し、それら 5 回の結果の平均をプロットしている。

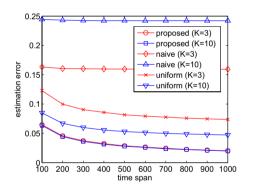


図 1: ノードの反多数派傾向の推定誤差

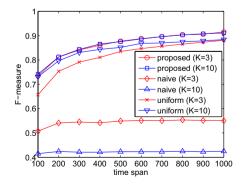


図 2: 反多数派傾向の高いノード群の抽出精度

期待通り、観測期間 T が増加する (すなわち、訓練データ \mathcal{D}_T の量が増える) につれて、推定誤差 \mathcal{E} は減少し F 値 \mathcal{F} は

増加することが見て取れる。そして、提案法が最も精度が高く、一様価値法が次に高く、ナイーブ法はかなり精度が低い。これらの結果は、我々が実験に用いたすべてのネットワークにおいて共通であった。また、すべての場合に対して提案法は、T=1000において 90% 以上の精度で反多数派オピニオニストを抽出することができた。さらに提案法は、一様価値法やナイーブ法と異なり、抽出精度がオピニオン数やネットワーク構造に依存して大きく変化しなかった。これらの結果は、提案法の有効性を実証しているとともに、適切なモデルを用いて最適化問題を解くことなしに、反多数派傾向の高いノードを抽出することは困難であることを示している。

また、提案法は、オピニオン価値も高精度に推定していたことに注意する。実際、オピニオン価値の平均推定誤差は、すべての場合において T=1000 で 1% 未満であった。さらに、提案法は、パラメータである反多数派傾向とオピニオン価値を効率良く推定していたことにも注意する。実際、提案法の T=1000 での処理時間は、K=3 のとき 3 分未満であり、K=10 のとき 4 分未満であった。ただし、これらの実験結果はすべて、Intel Core 2 Duo 3GHz プロセッサーの 1 台の PC で行ったものである。

6. おわりに

異なる価値を持つ複数のオピニオンが、反多数派オピニオニストの存在の下で、ソーシャルネットワーク上をどのように広がるのかという問題を、価値重みつき投票者モデルと反投票者モデルを結合した、価値重みつき混合投票者 (VwMV) モデルに基づいて研究した。VwMV モデルでは反多数派傾向を、2つのモデルの重みによって定量化し、パラメータとして扱った。我々は、本モデルを機械学習の枠組みで定式化し、各ノードの反多数派傾向と各オピニオンのオピニオン価値を、観測されたオピニオン拡散系列から最尤法により推定した。

提案パラメータ学習アルゴリズムは、十分な量のデータが存在すれば、反多数派傾向とオピニオン価値を効率良く高精度に同定できるということを、実ソーシャルネットワークトポロジーを用いた実験で実証した。特に、ノードの反多数派傾向を、そのノードの選択したオピニオンが近傍における少数派オピニオンであった回数を単純に数えることにより推定するという、ナイーブ法と比較し、提案法はそれに比べてはるかに精度が高いことを示した。

オピニオンシェアは各ノードの反多数派傾向に大きく依存し、それを高精度に推定できることは重要である。提案アルゴリズムによって学習したモデルは、将来のオピニオンシェア予測に利用できるとともに、様々な分析をするためのツールをも提供しうる。理論的解析により、VwMV モデルでは、近傍での期待局所オピニオンシェアがネットワーク全体での期待オピニオンシェアによって近似できるならば、オピニオン価値がポー様の場合には価値重みつき投票者モデルと異なり、オピニオン価値が最大のオピニオンが多数派になるとは限らないということを示し、オピニオン価値が一様な場合には、価値重みつき投票者モデルと同様に、オピニオンシェア予測問題は意味をなさないということ、すなわち、任意のオピニオンが多数派になりうるということを示した。また、これらの結果を、実ソーシャルネットワークトポロジーを用いたシミュレーション実験でも確認した。

謝辞:

本研究は、科学研究費補助金基盤研究 (C) (No. 23500194) の補助を受けた。

参考文献

- [Chen 09] Chen, W., Wang, Y., and Yang, S.: Efficient influence maximization in social networks, In KDD 2009, 199–208, 2009.
- [Domingos 05] Domingos, P.: Mining social networks for viral marketing, *IEEE Intell. Syst.*, 20, 80–82, 2005.
- [Even-Dar 07] Even-Dar, E., and Shapira, A.: A note on maximizing the spread of influence in social networks, In *WINE 2007*, 281–286, 2007.
- [Gruhl 04] Gruhl, D., Guha, R., Liben-Nowell, D., and Tomkins, A.: Information diffusion through blogspace, *SIGKDD Expl.* 6:43–52, 2004.
- [Kempe 03] Kempe, D., Kleinberg, J., and Tardos, E.: Maximizing the spread of influence through a social network, In KDD 2003, 137–146, 2003.
- [Kimura 10a] Kimura, M., Saito, K., Nakano, R., and Motoda, H.: Extracting influential nodes on a social network for information diffusion, *Data Min. Knowl. Disc.*, 20, 70–97, 2010.
- [Kimura 10b] Kimura, M., Saito, K., Ohara, K., and Motoda, H.: Learning to predict opinion share in social networks, In AAAI-10, 1364–1370, 2010.
- [Kimura 09] Kimura, M., Saito, K., and Motoda, H.: Blocking links to minimize contamination spread in a social network, ACM Trans. Knowl. Discov. Data, 3, Article 9, 2009.
- [Leskovec 07a] Leskovec, J., Adamic, L. A., and Huberman, B. A.: The dynamics of viral marketing, ACM Trans. Web, 1, Article 5, 2007.
- [Leskovec 07b] Leskovec, J., Krause, A., Guestrin, C., Faloutsos, C., VanBriesen, J., and Glance, N.: Costeffective outbreak detection in networks, In KDD 2007, 420–429, 2007.
- [Newman 02] Newman, M. E. J., Forrest, S. and Balthrop, J.: Email networks and the spread of computer viruses, *Phys. Rev. E*, 66, Article 035101, 2002.
- [Röllin 07] Röllin, A.: Translated poisson approximation using exchangeable pair couplings, *Ann. Appl. Prob.*, 17, 1596–1614, 2007.
- [Sood 05] Sood, V., and Redner, S: Voter model on heterogeneous graphs, *Phys. Rev. Lett.*, 94, Article 178701, 2005.
- [Wu 08] Wu, F., and Huberman, B. A.: How public opinion forms, In *WINE 2008*, 334–341, 2008.
- [Yang 09] Yang, H., Wu, Z., Zhou, C., Zhou, T., and Wang, B.: Effects of social diversity on the emergence of global consensus in opinion dynamics, *Phys. Rev. E*, 80, Article 046108, 2009.