

ネットニュースにおける対立の分析手法の提案

○瀬尾 雄三, 矢野 正晴

A Proposal of Method to Analyze Conflicts on the Net News

○Yuzo SEO, Masaharu YANO

Abstract

In the Net News, conflicts among the participants are often observed. It seems that the conflicts are often caused by the difference in the cultural background of the participants. In this paper, we intend to propose a quantitative extraction method of the conflicts. Articles in the Net News can be positioned in tree structures (threads) defined by the referential relation. A thread is treated as a set of samples for the analysis. Each article is distinguished by key word frequencies and they are summarized to principal component scores. Applying this technique to actual Net News threads, we found that when a thread includes conflicts, some scores vibrate with wavelength of two generations in the thread. Also the key of the conflict can be estimated with the meaning of the vibrating principal component. We expect that such a quantitative extraction technique of conflicts is useful for study of the social relation through electronic communication.

1. はじめに

ネットニュースでは、他人の依頼に応えた情報提供等の協調関係が随所にみられる一方、参加者相互の理解がなされず、対立だけが深まる議論もしばしば見受けられる。紛糾した議論の内容からその原因を推測すると、感情的行き違いもしばしば認められるものの、参加者間の文化的背景や常識の相違によると考えられる対立も多く認められる。ネットニュース上の議論における対立の分析は、電子的コミュニケーションを介した社会関係を解明する糸口になるのではないかと期待される。

電子的コミュニケーションにおける人間関係は、議論の内容を読むことでも分析可能であるが、膨大な記事を読解することは容易でない上、内容の読解には分析者の主観が作用する。そこで、これを機械的に解析する手段について検討を行った。

多変量解析を利用したメッセージ分析手法は古くから知られており(1)、メッセージを特徴に応じて配置して検索を容易にする手法(2)や、電子的メッセージの分類への利用(3)等が提案されている。これらは、メッセージの静的指標の抽出を目的とするものであるが、我々は、議論の過程での特性指標の変動を観測することにより、議論に含まれる対立を検出する可能性に着目した。

一連の議論を構成する記事の特性指標のスレッド内での変動は、以下のように、議論のあり方に関連すると考えられる。

- (1) 特性指標の長期的変化：議論の経過に伴って、中心となる話題が推移し、使われているキーワードも変化した場合に生ずると考えられる。
- (2) 枝分かれした議論間での特性指標の相違：元は一つであった話題が、議論の進行に伴い、異なる話題を扱う複数の議論に枝分かれした場合、それぞれの議論の枝では異なるカテゴリーのキーワードが用いられる可能性が高い。
- (3) 特性指標の短期的振動：異なるカテゴリーの用語が交互に用いられており、関心、常識、文化、思想等を異にする立場間での対立が生じている可能性がある。

対立が生じている場合、特性指標の異なるメッセージが交互に現れ、特性指標に短期的振動が現れる可能性が高いと期待される。この仮説が正しければ、特性指標の変動から短周期振動のみを抽出して定量化することにより、対立を検出することができるはずである。本報告では、この着想に基づく機械的分析手法と、実際のネットニュース上の議論にこれを適用して、対立の検出を試みた結果を紹介する。

2. 分析の手法

ネットニュースの記事は、相互の参照関係を識別するヘッダー情報を利用して、全ての記事を、参照関係で結合された、木構造にまとめることができる。今回の分析では、各木構造を、特定の話題に関する一連の議論（スレッド）とみなし、それぞれの木構造毎に、含まれる全ての記事をサンプルとして分析を行った。

キーワードには、スレッド内で出現頻度の高い、漢字の1以上の連なりを選択した。記事のキーワード出現度数をスレッド単位で主成分分析し、各記事の主成分スコアを求め、固有値上位の主成分スコアを記事の特性指標とした。

記事の特性指標の短い周期の振動は、議論を構成する一連の記事間での、特性指標の自己相関係数およびパワースペクトルにより評価した。枝分かれの多い木構造に配慮して、1～8世代隔てたスレッド間の全ての記事の組み合わせについて、主成分スコアの共分散を求め、これを自己相関関数として使用した。

3. 分析結果

3.1 分析対象

分析対象とするニュースグループには、ネットニュースの使い方について議論する場であり日常的に対立が観察される、fj.news.usageを選んだ。解析期間は1995年初めより1998年末までの4年間とした。この期間の記事総数は、重複記事を除き、50,294件であった。

多くの記事から構成されるスレッドは、対立関係を含む可能性が高いと思われることから、記事数で上位のスレッドを分析対象とした。

3.2 スレッド3の分析結果

記事数が3番目に多かった(576)スレッド(以下、スレッド3と呼ぶことにする)は「他人にはもっと親切に教えるべき」とする問題提起に始まり、ネットニュースという場の性格、他者との関係のあり方、妥当な言葉づかいなどに議論が進展した。

このスレッドの主成分分析結果を上位4主成分について表1に示す。第1主成分は記事の長さに対応する成分であった。第2主成分は論理や社会に関するキーワードが高い主成分負荷を示すのに対し、言語表現に関するキーワードが低い主成分負荷を示す。第3主成分は、「情報、記事」等のキーワードが高い主成分負荷を示すのに対して、「社会、一般」等のキーワードが低い主成分負荷を示し、提示される情報の内容と、情報交換で成り立つ人間関係のいずれを重視するかを表す主成分と考えられる。第4主成分は、「判断、自分」等のキーワードが高い主成分負荷を示す一方、「情報」等のキーワードが低い主成分負荷を示しており、主観を尊重するか、客観を尊重するかを表す成分と解せられる。

スレッド3の各スコアの自己相関係数を図1に示す。自己相関係数のラグ1の値は、フォロー記事とフォロー対象記事の主成分スコアの相関係数に相当するが、いずれも正の値を示し、フォローアップに際してはフォローの対象とされた記事と類似した用語が用いられる傾向があることを示している。

表1. スレッド3の主成分分析結果と特徴的キーワード(KW)

第1主成分	固有値 21.6 累積寄与率 15.0% 主成分の意味 記事の長さ
高負荷KW	思、人、言、方、的、私、出、一、対、事、考、全、実、持、同、
低負荷KW	初心者、最、所、常識、目、認、字、難、質問、初、馬鹿、失、動、化、由、使、
第2主成分	固有値 5.8 累積寄与率 19.0% 主成分の意味 論理：表現
高負荷KW	行、論、的、社会、立、数、明、理、実、一般、定、情報、主、同、上、
低負荷KW	関西弁、語、方言、礼、失、使、言葉、準、言、日、本、関、回、例、相手、良、
第3主成分	固有値 4.4 累積寄与率 22.0% 主成分の意味 情報：人間関係
高負荷KW	情報、記事、以、付、無、化、行、読、投稿、字、認、度、少、上、他、
低負荷KW	社会、一般、明、立、論、動、難、通、非、関係、一、理、同、全、用、当、
第4主成分	固有値 4.0 累積寄与率 24.8% 主成分の意味 客観性：主観性
高負荷KW	判断、自分、采、基、発言、不、人、手、意見、正、聞、考、世、誤、結、
低負荷KW	情報、例、化、対、関西弁、同、実、行、礼、語、失、付、的、字、方言、以、

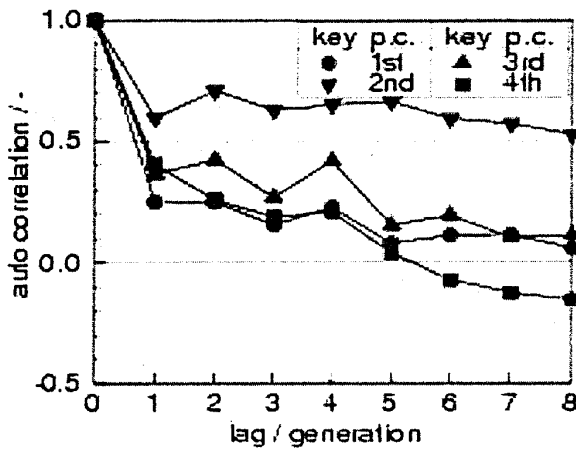


図 1.スレッド3の自己相関係数

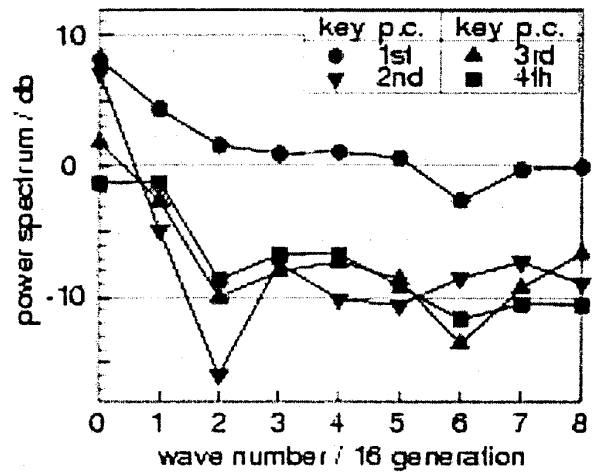


図 2. スレッド3のパワースペクトル

第3主成分の自己相関係数は、ラグ2を周期とする波打ちを示す。これは、議論の過程で第3主成分スコアの高い記事と低い記事とが交互に現れることを示し、対立の存在を示唆する。これをフーリエ変換して得たパワースペクトルを図2に示す。横軸は16世代あたりの波数を単位とし、数8は2世代周期を意味する。第3主成分のパワースペクトルは、2世代周期で増加を示し、自己相関係数のラグ2を周期とする波打ちに対応している。

第3主成分スコアの高い記事は、伝達される情報を重視する記事であり、第3主成分スコアの低い記事は、伝達行為自体の妥当性を重視する記事と考えられる。第3主成分スコアの2世代を周期とする変動は、情報伝達に注目する主張と、社会常識に注目する主張が交互になされていることを示し、双方の立場間で対立が生じていることを示唆する。

このスレッドに含まれる記事を読解した結果、他者との良好な関係を重視するか、正確な情報を重視するかという対立関係が認められた。他者との良好な関係を重視する論拠は社会常識に求められており、これは、振動が認められた主成分の意味とよく対応している。

3.3 大きなスレッドの分析結果

分析対象範囲において、構成記事数の多い上位20スレッドとこれに対する分析結果を表2に示す。振動の測定は、固有値が1以上である全ての主成分について行った。表中の振動強度は2世代周期におけるパワースペクトルの値(ピークとベースラインの比)の最大値をdB単位で示したものであり、これが5dB以上のスレッドは、ほぼ対立が含まれることが読

表 2. 20 スレッドの対立の分析結果

順位	大 きさ	振動 強度	語句 の数	主成 分数	分類(右表参照)					内容
					1	2	3	4	5	
1	844	2.85	150	46					x	署名の引用
2	673	3.47	150	46			x	x		匿名投稿, end行
3	576	8.97	144	42	x					初心者です
4	574	6.99	143	42		x				用語(レス, 他)
5	480	8.77	120	33			x			end test
6	456	5.61	114	34	x		x			end行の理由
7	455	6.29	113	33	x					初心者より一言
8	412	3.53	103	28	x					厳しい指摘
9	374	9.24	93	29	x					厳しい指摘
10	367	6.89	91	25	x	x			x	議論の仕方
11	364	3.36	91	29	x	x				厳しい指摘, 文字コード
12	357	6.65	89	26	x					注意の仕方
13	343	8.38	85	26		x				差別用語
14	338	4.32	84	25	x				x	初心者用ニュースグループ
15	337	3.56	84	26			x			end行の文字化け
16	328	24.52	82	23	x					指摘の仕方
17	289	7.03	72	19	x	x				文字コード, 指摘の仕方
18	286	5.76	71	18	x				x	文書の多量転載
19	285	3.67	71	21	x				x	fjと組織, メール転載
20	273	5.38	68	17	x					批判の仕方

注1) 太字は振動が認められたスレッドを示す

注2) 各分類番号の内容は表3に示す

表 3. 議論の分類

分類	議論の内容	対立概念
1	指摘の仕方	親密性: 普遍性
2	用語について	通用性: 正確性
3	end行について	通用性: 普遍性
4	匿名投稿について	
5	ネットニュースの利用法	

解によって確認された。

このニュースグループの大きなスレッドで扱われた話題として、表2の分類の欄に示すように、特に初心者に対する指摘のし方、用語を正確に用いるべきか否か、バグのあるニュースリーダーに異常動作を起こさせる end 行を記事に含めることの是非の3つの話題が大きな議論を呼んでいた。これらの議論で見られる対立は、図3に示すように、正確な情報、表現を重視し、規約の遵守を求める立場と、温かみのある人間関係と実用性を重視する立場の間の対立に集約することができる。前者は古くからの利用者である学術関係者の普遍性指向と、後者は近年急増したインターネット利用者のコミュニティー指向と、それぞれ重畳するように思われる。

4. 考察

ネットニュース上の大きなスレッドにここで提案する手法を適用した結果、一部の主成分スコアに2世代を周期とする振動が存在すると判断され、内容から解釈される対立点との対応が認められた。同様な結果は他のスレッドでも確認されている(4)。

大きなスレッドは、ほぼ例外なく何らかの対立する主張を含んでいるが、双方が他方の主張を尊重する形で議論する場合は、主成分スコアの振動が観察されない。これは双方のメッセージに、類似したキーワードが現れる傾向が高いためと思われる。一方、他者の主張に対する理解を欠いた主張が繰り返される場合は、主成分スコアに振動が認められる。即ち、この手法で抽出される対立は、単なる主張の対立ではなく、相互理解不能という、より深刻な対立関係と考えられる。これは、社会に対する常識、ないし、文化の対立とも言えよう。

図4は、ネットニュースで認められた対立点を、社会的指向性と関連付けたものである。電子的コミュニケーションは、親密な人間関係に支えられた閉鎖的で小さな社会を形成する一方で、匿名的人間関係の支配する開かれた大きな社会を形成するという、矛盾した機能を併せ持ち、同一の場に文化・嗜好の異なる二つのグループ形成を促す作用がある。これがコミュニケーションの場における対立とも符合していることは興味深い、これら社会的・文化的側面を掘り下げる議論は今後の研究に委ねたい。

参考文献

- 1) Krippendorff, K., "Content Analysis: An Introduction to Its Methodology," Sage, Beverly Hills, CA., 1980. (三上俊治, 椎野信雄, 橋本良明訳「メッセージ分析の技法—内容分析への招待」, 勁草書房, 1989)
- 2) Deerwester, S. et al., "Indexing by Latent Semantic Analysis," *Journal of the American Society for Information Science*, Vol.41, No.6, pp.391-407, 1990.
- 3) 小川知也, 落谷亮, 西野文人, 「文書クラスタの判別のための特徴表現付与」, 言語処理学会第5年会年次大会発表論文集, pp.209-212, 1999.
- 4) 瀬尾雄三, 矢野正晴, 「ネットニュース参加者間の対立の定量的分析」, *NII Journal*, No. 1, pp.51-62, 2000

瀬尾 雄三 東京大学先端経済工学研究センター (〒153-8904 東京都目黒区駒場四丁目6番1号)
 矢野 正晴 国立情報学研究所 (〒101-8430 東京都千代田区一ツ橋二丁目1番2号)
 Yuzo SEO (seo@rc.m.kagaku.co.jp) Research Center for Advanced Economic Engineering, The University of Tokyo
 Masaharu YANO (yano@nii.ac.jp) National Institute of Informatics

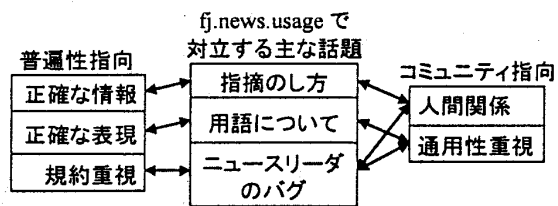


図3. 主な対立点

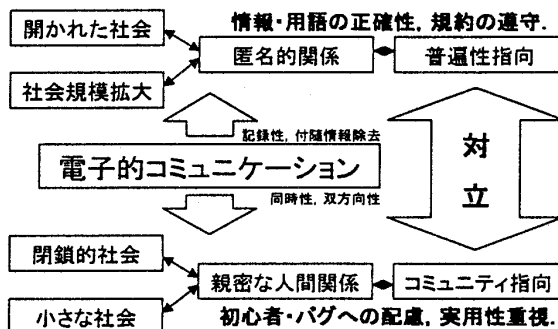


図4. 対立の社会的背景