

研究論文

ネットニュース参加者間の対立の定量的分析

Quantitative Analysis of Conflicts among Participants in the Net News

瀬尾 雄三

東京大学先端経済工学研究センター

Yuzo SEO

Research Center for Advanced Economic Engineering, The University of Tokyo

矢野 正晴

国立情報学研究所

Masaharu YANO

National Institute of Informatics

要旨

インターネット上の掲示板サービスであるネットニュースの参加者は、組織の枠を越えて見知らぬ他者との間でさまざまな話題について議論を交わしている。そこでは、参加者間の対立がしばしば観測されるが、そういった対立が参加者達の文化的背景が異なることによるものであることが多くあるように思われる。一方、ネットニュースにおいて一連の議論（以下、スレッドと呼ぶ）を構成する個々の記事は、記事相互間の参照関係によって木構造に関係づけることができ、また、各々の記事は、その中に含まれるキーワードの出現頻度により特徴づけることができる。そして、キーワードの出現頻度は、主成分スコアに要約される。そこで、我々は、主成分スコアをもとに、自己相関係数とパワースペクトルという二つの特性指標を算出し、それらによりネットニュース参加者間の対立を定量的に抽出することを試みた。ネットニュース上の4つのスレッドにこの手法を適用したところ、2世代を周期とする特性指標の振動が広範に認められ、振動の観測された特性指標の意味と、議論内容を読解することにより得られる対立点とを比較した結果、文化的背景を異にする参加者間の議論において、双方が一致する傾向が認められた。このような対立の定量的抽出手法は、電子的コミュニケーションを介した社会関係を研究する上で有益と考えられる。

ABSTRACT

The participants in the Net News, services on the Internet, conduct debates about the various topics with the strangers beyond the frame of organizations. There conflicts among the participants are often observed, but it seems that there are a lot of cases that those conflicts are caused by the difference in the cultural backgrounds of the participants. On the other hand, a series of articles (hereinafter called a thread) in the Net News, can be related to the tree structure by the referential relation among the articles. Each article can be distinguished by the frequency of key words contained in it, and the frequency of the key words is summarized to the principal component scores. Therefore, we tried to calculate two characteristic indexes, the auto-correlation coefficient and the power spectrum, based on the principal component scores, and to extract conflicts among the Net News participants quantitatively. Applying this technique to four threads on the Net News, the vibration of the characteristic indexes, the period of which was 2 generations, was extensively observed. In comparing the meaning of the characteristic indexes in which the vibration is observed and the opposing point which is gotten by reading the thread contents, there was the tendency that both coincide with each other in the case of argument among the participants who have different cultural backgrounds. It seems that such a quantitative extraction technique of conflicts is useful when studying the social relation through electronic communication.

[キーワード]

ネットニュース、スレッド、対立、主成分分析、自己相関係数、パワースペクトル

[Keywords]

Net News, thread, conflict, principal component analysis, auto-correlation coefficient, power spectrum

1. はじめに

新しいコミュニケーション技術は、人々の関係のあり方を変え、社会のあり方にも変化を及ぼすと言われている (McLuhan(1964)[1])。近年急速に利用が拡大している電子的コミュニケーション技術も、我々の社会関係に大きな変化をもたらす可能性が高く、その影響を予測することは、新しい技術を効果的に活用する上で欠かせない。

電子的コミュニケーションが極限にまで普及した社会は、電子的コミュニケーションを介して成立している社会として、既に小規模ながら存在している。すなわち、電子メールを主な連絡手段とする人々の集団や、電子掲示板で結び合わされた人々の社会は、将来の社会関係を先取りした社会とみなすこともでき、これらの限定的な社会関係を分析することで、一般社会の将来像を描き出し、いずれ遭遇する可能性のある問題を早期に見出すことができると期待される。

インターネット上の公開された電子的コミュニケーションの場であるネットニュースにおいても、見知らぬ他者の質問に応えた情報提供等の協調関係が多くみられる一方で、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。電子的コミュニケーションは、伝統的なコミュニケーションに比べて心理的・社会的制約が低く、よりフランクな議論がなされる一方で、表情や声のトーン等の副次的な情報が伝達されず、相手の感情を発信者にフィードバックする機能に乏しいために、参加者間の争いも招きやすいと言われている (Sproull(1992)[2]) が、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、より深い原因は参加者間の文化的背景や常識の相違にあると考えられるケースが多いものと思われる。このことは、議論における対立を分析することで、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。

対立の分析は、一般的には議論の内容を読むことによってなされる。例えば、Johnston 他 (1999)[3] は掲示板上の議論の読解により、集産主義的文化と個人主義的文化の対立を指摘している。しかしながら、ネットニュースに投稿される膨大な記事を読解することは容易でなく、また、内容の読解による対立の検出には、分析者の主観が作用するという問題がある。

そこで、電子的コミュニケーション上の社会関係を客観的に解析する手段として、対立の定量的評価手法に

ついて検討し、ネットニュースに投稿された実際の記事の分析を行い、評価手法の妥当性を検討した。

2. 先行研究との関連

電子的コミュニケーションの分野では、メッセージの機械処理が容易であることから、さまざまな解析手法が提案されている。参加者相互の関係に注目した研究として、個々のメッセージに対するフォローアップの数を解析してコミュニケーションの場の特性を把握する試み (川上他 (1993)[4]、瀬尾 (1996)[5]) がなされているが、これらはメッセージの数のみを扱い、内容には触れられていない。

議論の推移の把握を目指した研究として、電子メールを介して行われる協同作業のメッセージから、問い合わせと返答のペアを機械的に抽出し、会話の完結度を解析したもの (村越・落水 (1998)[6]、村越・島津・落水 (1998)[7])、話題の推移に関連した重要メッセージの抽出 (井佐原他 (1997)[8]、内元他 [9](1997)、内元他 (1998)[10])、話題の消長の自動検出 (齊藤他 (1998)[11]) などが研究されており、記事の内容から議論のダイナミズムを解析する試みとして注目される。これらの手法は、特定のパターンにマッチするメッセージを検索しており、特定の技術やシステムの検索等を想定した手法といえる。

特定の目的を持ったメッセージの検索には少数のキーワードが十分な手がかりになると考えられるが、通常、ネットニュースに投稿される記事には多種多様な一般的な用語が用いられており、その解析に際してはより抽象性の高い概念にキーワードを統合する必要があると考えられる。

多数のキーワードを少数の指標に統合するメッセージ内容の分析手法として、多変量解析に基づく手法が古くから知られている (Krippendorff(1980)[12])。これらの手法においては、多数のメッセージを含む母集団を対象に、個々のメッセージにおけるキーワード出現頻度を因子分析、もしくは主成分分析し、各々のメッセージのスコアをメッセージの特性を代表する指標とみなす。

多変量解析を利用したメッセージ分析手法の具体例として、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 (Deerwester(1990)[13]) や、電子的メッセージの分類への利用 (小川他 (1999)[14]) が提案されている。これらはいずれも静的な指標の抽出に止まっているが、議論の過程におけるメッセージ特性指標の変動

から、議論そのものの特性も把握できる可能性がある。

対立を抽出するという目的から、一連の議論を構成するメッセージを母集団として分析すると、その内部における特性指標は、以下の、議論の過程に含まれる要因によって変動するものと考えられる。

- (1) 時間の経過に伴う話題の変化
- (2) 異なる話題を扱う枝分れした議論
- (3) 論点、立場を異にする参加者間での論戦

上記最後の要因は対立に対応する。すなわち、二つの立場の間で対立が生じている場合、特性指標が異なるメッセージが短い周期で交互に現れる確率が高いことが期待される。この仮説が正しければ、議論の過程における特性指標の短周期の変動を測定することにより、対立を検出することができるはずである。そこで、これを定量的に検出する手法を開発し、実際のネットニュース上の議論に対してこの手法を適用して対立を検出することを試みた。

3. 分析の手法

3.1 分析対象テキストの抽出

図1にネットニュースの記事の構成を示す。記事はヘッダー部と本文で構成され、ヘッダー部の最後は空行で識別される。ヘッダー部の記述はRFC-1036[15]で定められた様式に従い、投稿者のメールアドレス、メッセージID、投稿年月日、および、他の記事に対する応答記事である場合は応答先のメッセージIDその他が記述されている。

筆者らは、ヘッダー部に続く記事本文からキーワードの出現頻度を計測した。また、本文中で他の記事を引

Path: 記事の転送経路
From: 投稿者のメールアドレス
Newsgroups: ニュースグループ
Subject: サブジェクト (表題)
Message-ID: <メッセージID>
Date: 投稿年月日
References: <応答先メッセージID>
記事本文
署名

図1: ネットニュース記事の構成

用している部分は、行頭文字により識別し、計測対象から除外した。ハイフン2文字もしくは3文字のみからなる行以降はシグネチャとみなした。さらに、定型的挨拶部分を削除するため、同一投稿者の記事の20%以上かつ3件以上が先頭または末尾に同一の文字列を含む場合、一致部分を削除した。この操作により、全文が削除される記事も存在したが、これらは定期投稿文書および誤操作によって多数回投稿されたと思われる記事であり、解析対象から除外して妥当と考えた。

3.2 キーワードの検出

キーワードには漢字の1以上の連なりを採用し、以下の手順にしたがって機械的に抽出した。抽出するキーワードの数は、スレッドを構成する記事数の1/4とし、これが150以上となる場合は150とした。

- (1) キーワードリストの初期値として、スレッド内で出現度数の高い単漢字を、十分な数(抽出キーワード数の2倍とした)だけリストに登録する。
- (2) 解析対象テキストをスキャンして、リストに登録されたキーワードの出現度数とキーワードの連なりの出現度数を計数する。
- (3) キーワードとその連なりを一括して出現度数でソートし、出現度数が上位の連なりについては、これを構成する二つのキーワードを接続してキーワードリストに加えるとともに、出現度数が下位の分解可能なキーワードについては、これを分解してキーワードリストより除去する。
- (4) (2)、(3)の操作を、100回を上限として、変化がある間繰り返す。

手順(3)においてキーワードの競合を避けるため、同一のキーワードが出現度数上位の二つ以上の連なりの前後に同時に含まれる場合、出現度数が最も高い連なりのみを接続操作の対象とし、他の連なりに対する操作は次回の繰り返しに持ち越した。なお、同一の連なりについて接続・分解が繰り返されることを防ぐため、接続操作は出現度数の順位が抽出キーワード数よりも2以上上位の連なりを対象とし、分解操作は出現度数の順位が抽出キーワード数よりも2以上下位のキーワードを対象とした。

3.3 スレッドの形成

この解析はスレッドを対象に行った。ネットニュースに投稿される記事は、新規の話題を提供する記事と既に投稿された他の記事をフォローアップする記事とに大別される。フォローアップ記事はヘッダ部に **Reference** 行を含み、この部分にフォローアップの対象となる記事のメッセージ ID が記載されている。この情報を用いることで、記事を参照関係に従う木構造に関係づけることができる。こうして形成された一つの木構造を一つのスレッドとみなして、分析の対象とした。

フォローアップの対象とされた記事が転送の過程で失われた場合、あるいは、それが分析対象範囲外の記事（他のニュースグループに投稿された記事や、分析期間以前に投稿された記事）であった場合、その部分で木構造が分断されるという問題を生じる。この問題を回避するため、**Reference** 行の先頭に書かれたメッセージ ID に該当する記事が分析対象範囲に存在しない場合、これをメッセージ ID とするダミー記事を挿入した。フォローアップが続けて行われる場合、**Reference** 行は複数のメッセージ ID を含む場合が多い。ダミー記事の **Reference** 行を、元の **Reference** 行から先頭のメッセージ ID を除いたものとして、上記操作を繰り返すことで、可能な限りの木構造の回復を図った。

今回の分析範囲において、スレッド総数は 5,998 であり、そのうち二つ以上の記事から構成されるスレッドは 3,197 であった。また、最終的に挿入されたダミー記事は 2,340 件であり、ダミー記事を含む解析対象記事総数は 52,634 件であった。

3.4 主成分分析

それぞれのスレッドにつき、スレッドを構成する記事をサンプルとし、キーワードの出現度数を変数とする主成分分析を行い、各記事の主成分スコアを求めた。分析は固有値の大きい上位 4 主成分について行った。各主成分の意味は、キーワードを主成分負荷でソートして高い負荷を示す 15 キーワードと低い負荷を示す 15 キーワードを抽出すると共に、記事を主成分スコアでソートして主成分スコアの低い 10 記事と主成分スコアの低い 10 記事を抽出し、これらの内容を吟味することにより判定した。

分析には、それぞれの記事に含まれるキーワードの出現度数を用いた。その結果、全ての分析において第 1 主成分の寄与率が際だって高い値を示した。表 1 に示すように、第 1 主成分スコアの低い記事は本体行数が少

表 1：記事の本体行数と第 1 主成分スコア順位

スレッド	スコア上位記事の本体行数			スコア下位記事の本体行数		
	1 位	2 位	3 位	1 位	2 位	3 位
1	143	120	146	6	12	14
3	229	195	133	11	27	12
4	179	89	103	5	12	6
6	158	96	131	9	61	9

なく、この成分は大きさの因子と考えられる。主成分分析において大きさの因子が第 1 主成分に現れることは珍しくないが、ネットニュースでは、特に、長い記事と短い記事が混在しているため、大きさの因子が際だって高い寄与率を示したものと考えられる。第 5 章でスレッドごとに主成分分析の結果を説明するが、ここでは第 1 主成分についての説明を省略する。

記事の長さの影響を除去するため、漢字数に対するキーワードの出現度数の比率を用いた主成分分析も試みた。この場合、大きさの因子が高い寄与率を示す現象は回避されるが、キーワードをほとんど含まない短い記事の主成分スコアが異常値を示す例が多く、ノイズが増大することから、比率を用いた解析はあまり適当でない判断した。なお、度数を用いた分析にも、大きさの因子が変動した場合、他の全ての主成分スコアも変動するという問題が残っている。

3.5 自己相関係数およびパワースペクトルの算出

本稿で提案する手法は、議論の過程における主成分スコアの変動から、議論に含まれる対立を検出するという着想に基づいている。変化する現象を分析する手法としてはスペクトル解析が一般的であるが、分枝が多数存在するネットニュースのスレッドでは、十分な長さの変数系列が得られないという問題がある。

自己相関関数は、時間に関する不規則変量の τ 時間隔たった二つの変動の積のアンサンブル平均として定義されるが (日野 (1977)[16])、ラグ τ における関数値は τ 時間隔たった二つの変量間の共分散に他ならない。共分散は、統計的性質が一定とみなされる範囲内であれば、 τ 時間隔たった二つの変量の全ての組から計算可能であり、分枝の存在する系でも大きなサンプル数を用いて解析できる。この論文では、スレッド毎に 1～8 世代隔てた記事間のスコアの共分散を計算して自己相関関数を求め、これから自己相関係数とパワースペクトルを算出し、世代による振動の有無を調べた。

4. 分析対象

分析の対象とするニュースグループとして、日本国内を中心に活発な投稿が行われているfj.news.usageを選択した。このニュースグループはネットニュースの使い方について議論する場とされており、さまざまな対立が日常的に観察される。解析期間は1995年1月1日より1998年12月31日までの4年間とし、全てのデータは北陸先端科学技術大学院大学のfjアーカイブ[17]より入手した。重複記事を除いた記事総数は50,294件であった。

多くの記事を持つスレッドは、何らかの対立関係を含んでいる可能性が高いと思われることから、大きなスレッドを分析の対象にすることとした。分析対象範囲において400件以上の記事を含むスレッドは表2に示す8スレッドであった。これら8スレッドに対して本手法を用いて特性指標の振動検出を試みた結果、同表に"+"を記した5スレッドに振動が検出された。スレッド1には振動が認められず、スレッド2及び8についてはノイズと振動の判別が困難であった。

以下、上記の8つの大きなスレッドの中から、紙幅の都合で4つを選び、分析結果を紹介するとともに、振動の認められた主成分の意味と読解によって認められたスレッドに含まれる対立とを比較し、手法の妥当性を検討する。

5. 分析結果

5.1 スレッド1(署名の引用)

スレッド1は、「Quotation of signature」と題する記事に始まる、分析対象内で最大のスレッドである。この

表2: 大きなスレッドとその解析結果

順位	記事数	振動解析結果	内容
1	844	-	署名を引用部を含めることの是非
2	673	?	偽名投稿、ニュースリーダーとend行
3	576	+	初心者への対応のあり方
4	574	+	「レス」等の用語について
5	480	+	end行対策としての全文引用の是非
6	456	+	end行を入れる理由とその是非
7	455	+	初心者への対応のあり方
8	412	?	指摘の仕方

スレッドは、フォローアップに際して自分の記事の署名部分まで引用された者が不快感を表明したことに始まり、正当な引用の方法、fjにおけるマナー、あるいは他人を「無神経」と評することの是非など、さまざまな点について議論が行われた。

このスレッドの主成分分析結果を表3に示す。第2主成分では、「無神経」、「嫌」といった感性的表現に用いられるキーワードが高い主成分負荷を示す一方、主成分負荷の低いキーワードに著作権法に関する用語が並ぶことから、感性対知性の成分と解釈できる。第3主成分では、「可能性」、「理由」等の議論の内容を問うキーワードが高い負荷を示す一方で、「質問」等、議論の方法に関わるキーワードが低い負荷を示す。これらは、双方とも著作権法に関係するキーワードであるが、第4主成分スコアの高い記事には自らの行動を問う内容の記事が多く、スコアの低い記事には一般的事実を紹介する記事が多いという違いがある。

次に、スレッド1の主成分スコアの自己相関係数を図

表3: スレッド1(署名の引用)の主成分分析結果

第1主成分	固有値 20.0	累積寄与率 13.3%	主成分の意味 記事の長さ
負荷の高いキーワード	引用, 思, 私, 考, 一, 記事, 上, 方, 言, 場合, 人, 的, 明, 示, 書		
負荷の低いキーワード	日, 説, 僕, 誰, 存, 常, 理解, 単, 持, 無駄, 心, 通, 入, 可能性, 表現, 話		
第2主成分	固有値 6.6	累積寄与率 17.7%	主成分の意味 感性: 知性
負荷の高いキーワード	無神経, 嫌, 人, 気, 今, 私, 回, 数, 初, 知, 田, 不, 対, 論, 少		
負荷の低いキーワード	表示, 第, 十, 条, 著作者名, 用, 名, 著作者, 著作物, 場合, 示, 著作権法, 部, 名前, 権利, 定		
第3主成分	固有値 3.6	累積寄与率 20.1%	主成分の意味 議論の内容: 議論方法
負荷の高いキーワード	性, 範, 困, 引用, 可能性, 著作権法, 不, 主張, 関係, 例, 著作物, 当, 作, 理由, 著作権		
負荷の低いキーワード	著作者名, 質, 田, 中, 名, 表示, 問, 指, 変, 回, 記, 初, 時, 名前, 著作者, 私		
第4主成分	固有値 3.3	累積寄与率 22.3%	主成分の意味 主観性: 客観性
負荷の高いキーワード	範, 困, 田, 署名, 問, 野, 質, 中島, 心, 目, 初, 著作権法, 正, 明, 者		
負荷の低いキーワード	気, 付, 他人, 情, 無神経, 名, 著作物, 人, 発, 常, 嫌, 感, 表示, 日, 実, 不		

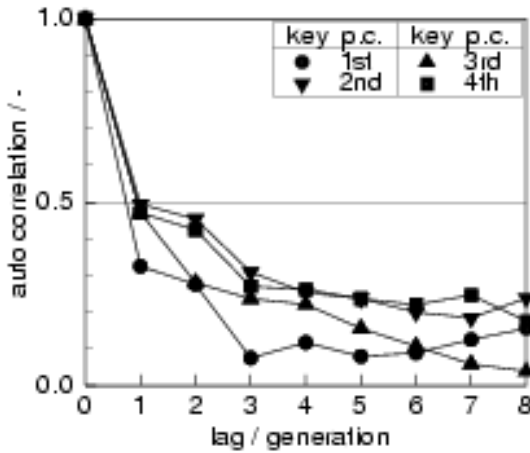


図2: スレッド1 (署名の引用) の自己相関係数

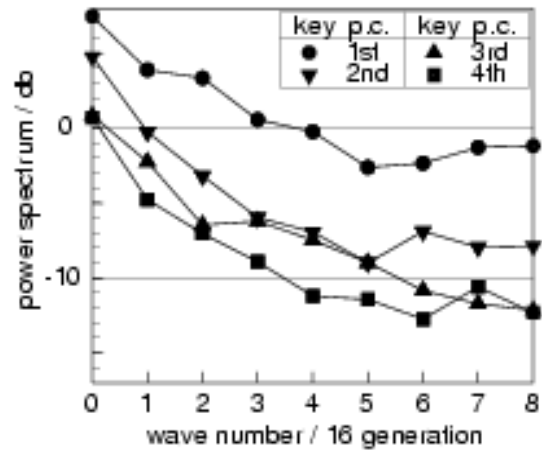


図3: スレッド1 (署名の引用) のパワースペクトル

2に、それをフーリエ変換して得たパワースペクトルを図3に示す。ラグ1の自己相関係数はすべての主成分で、本稿で分析する4つのスレッドの中で比較的高い値を示すが、ラグ数の増加に伴って減少しており、論点によるパスの分離は明瞭ではないと考えられる。このスレッドのラグ1の自己相関係数が高いという結果は、フォローアップがその対象とする記事と類似した用語を用いてなされている傾向を示し、互いに相手方の主張に配慮した協調的な議論がなされている可能性を示している。主成分スコアのパワースペクトルにも2世代を周期とするスペクトルの上昇は認められなかった。

一方、このスレッドに含まれる記事を読むと、「無神経」という用語に対する認識の相違など、いくつかの相反する主張が含まれるが、相互に相手の主張を理解する形で議論が行われており、深刻な対立は見出されなかった。これは、主成分スコアの分析結果と一致する。

5.2 スレッド3 (初心者への対応)

スレッド3は「初心者です」という表題で「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、言葉づかいなどに議論が発展した。

このスレッドの主成分分析結果を表4に示す。第2主成分は論理や社会に関係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し、論理性対表現方法の主成分と考えられる。第3主成分では、「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、第3主成分スコアの記事内容も参照して解釈すると、情報(コミュニケーションの内容)と社会関係(コミュニケーション行為そのもの)のいずれを重視するかを表す主成分と考えられる。第4主成分では、「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、客観性を尊重するか、主観性を尊重するかを表す成分と解せられる。

次に、スレッド3の各主成分スコアの自己相関係数を図4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。

第2主成分の自己相関係数は、他の主成分の自己相関係数と異なり、ラグ数の増加に伴う低下が少ない。これは、一つのスレッド中に第2主成分に関して性質の異なるサブツリーが存在していることを示唆する。第2主成分では、「関西弁」などの言語表現に関するキーワードのある記事が低い主成分負荷を示すが、最初に関西弁に言及した記事[18]以下のサブツリーとそれ以外にスレッドを分離し、それぞれに属する記事の各主成分スコアの平均値を求めたところ、表5に示すように、両者で第2主成分スコアに大きな差が認められた。自己相関係数が計算される議論のパスの大部分はサブツリーの内外いずれか一方の側にあるため、第2主成分スコアの自己相関係数は高く安定した値を示したものと考えられる。

表 4 : スレッド 3 (初心者への対応) の主成分分析結果

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
負荷の高いキーワード	思, 人, 言, 方, 的, 私, 出, 一, 対, 事, 考, 全, 実, 持, 同	
負荷の低いキーワード	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
負荷の高いキーワード	行, 論, 的, 社会, 立, 数, 明, 理, 実, 一般, 定, 情報, 主, 同, 上	
負荷の低いキーワード	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 : 社会関係
負荷の高いキーワード	情報, 記事, 以, 付, 無, 化, 行, 読, 投稿, 字, 認, 度, 少, 上, 他	
負荷の低いキーワード	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
負荷の高いキーワード	判断, 自分, 来, 基, 発言, 不, 入, 手, 意見, 正, 聞, 考, 世, 誤, 結	
負荷の低いキーワード	情報, 例, 化, 対, 関西弁, 同, 実, 行, 礼, 語, 失, 付, 的, 字, 方言, 以	

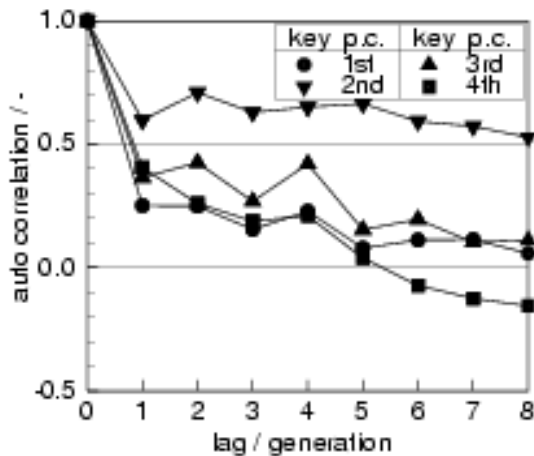


図 4: スレッド 3 (初心者への対応) の自己相関係数

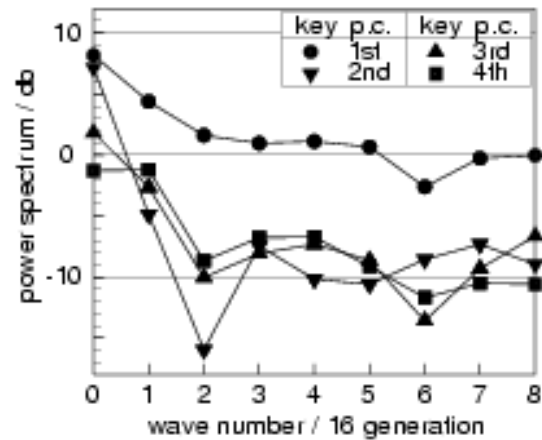


図 5: スレッド 3 (初心者への対応) のパワースペクトル

表 5 : スレッド 3 のサブツリー内外での主成分スコアの平均値 (かっこ内は標準偏差)

平均範囲	記事数	第 1 主成分スコア	第 2 主成分スコア	第 3 主成分スコア	第 4 主成分スコア
内部	176	3.9 (5.8)	-1.6 (3.5)	-0.1 (1.1)	0.0 (1.8)
外部	397	2.9 (4.2)	-0.1 (1.3)	0.3 (1.7)	0.1 (1.3)
全体	573	3.2 (4.8)	-0.5 (2.2)	0.2 (1.5)	0.0 (1.5)

一つのスレッドの内部に異なる性質のサブツリーが含まれる場合、これを分離して分析することも考えられるが、一連の議論がさまざまな論点を含むことは一般的であり、出現するキーワードが異なるだけではスレッドを分離して解析する理由にはならないと考え、振動の検出に関してはスレッドの分離は行わなかった。

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が交互に投稿されていること

を示し、対立の存在を示唆する。自己相関係数をフーリエ変換して得たパワースペクトルを図 5 に示す。図の横軸は 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 世代を周期とする点で増加しており、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、情報の観点 (コミュニケーションの内容) にたつ議論であり、第 3 主成分スコアの低い記事は、一般・社会という観点 (コミュニ

ケーション行為そのもの)にたつ議論である。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、一方を重視する立場と他方を重視する立場の間で対立が生じていることを示唆する。

一方、このスレッドに含まれる記事から読み取られたのは、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であった。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。

5.3 スレッド4(「レス」等の用語)

スレッド4は、哲学のニュースグループで行われていた議論の過程で、「ホームページ」「レス」等の用語を用いた者に対しその使い方は誤用であるとの指摘がなされていたが、そのニュースグループをfj.news.usageに移して、これらの用語を用いることの是非及びこれを指摘することの是非を巡って議論がなされたものである。用語に関する議論としては、解析対象範囲で最大のスレッドである。

このスレッドの主成分分析結果を表6に示す。また、自己相関係数を図6に、パワースペクトルを図7に示す。特に第3主成分の自己相関係数において振動が顕著に見られ、第3主成分のパワースペクトル、第2、第4主成分の自己相関係数とパワースペクトルにも若干の振動が見られる。

第3主成分スコアの高い記事と低い記事は、用語を正確に使うべきか否かで別れているわけではない。第3主成分スコアの高い記事は論理性と具体例に基づく議論であり、第3主成分スコアの低い記事には感性的な発言が目立つ。同一の主成分軸の、一方の側には用語を正確に使うべきとする主張に対する論理性の高い賛否両論が現れ、他方の側には感性的発言が現れたことは、誤用を指摘する側にも感性的要素が作用している可能性を示す。第3主成分に次いで大きな振動が観察された第4主成分に関しては、スコアの高い記事は、定着している誤用について「列車と電車」、「時間と時刻」などの例を挙げて説明するものが多く、第4主成分スコアの低い記事には、誤用を指摘する必然性を論じた記事が多い。

このように、このスレッドにおいては分析の結果、振動は検出されたものの、読解によって認められた対立点と主成分分析によって検出された振動成分の意味と

は一致していない。スレッドの中核をなす議論は、用語の誤用を指摘することは正当であるとする主張と、その場で理解できる程度の誤用であれば、これを指摘することはかえってコミュニケーションを阻害とする主張の対立であるが、双方とも相手方の主張を良く把握した上で議論を進めており、各々のメッセージが同種のキーワードを含むため、この手法では対立が検出されなかったものと考えられる。

振動が検出された第3主成分スコア、第4主成分スコアが特異な値を示した記事は、中核的なパスの周囲の枝に相当するパスに存在する。すなわち、このスレッドは、主張を異にするが知性的な論者による協調性の高い議論が中核をなす一方で、これに感性的な反応を示した短い枝がいくつか付随するという形態を持つためと考えられる。周囲の枝において議論のすれ違いが生じた理由として、「レス」等の用語をネットニュース上で用いることの妥当性に関しては以前より議論がなされており、これを擁護する記事や誤用を指摘する記事に対して、個々の実例や表現に触発された、感性的な記事が投稿されたのではないかと思われる。

5.4 スレッド6(end行の理由)

スレッド6は「end行の理由」という表題で開始された記事数456からなるスレッドである。ここで問題とされている「end行」とは、「end」のみからなる行を意味する。マイクロソフト社から配布されたニュースリーダ「MSIN」は、end行を含む日本語メッセージを表示する際に文字化けを起こすという問題があった。この現象に関しては、いくつかのスレッドで大きな議論に発展したが、このスレッドはその典型的な一例であり、投稿記事にend行を意図的に入れたと思われる投稿者に対する非難と、end行を記事に含めることは正当であるとする主張の間で議論が行われている。

このスレッドの主成分分析結果を表7に示す。第2主成分は、文字化けを指摘する方法に関する記事が高いスコアを示す一方で、感性的反応を内容とする記事が低いスコアを示す。第3主成分は、記事内容の読解も合わせて行い検討したところ、end行を付けるという行為の意味を断定的に論じる記事が高いスコアを示す一方、普遍性の高い規約を尊重してend行の善悪判断を留保する主張が低いスコアを示している。第4主成分は、「自分」、「勝手」等の語句を含む感性的記事が高いスコアを示す一方で、end行を付ける理由を述べた比較的知性的な記事が低いスコアを示している。なお、一

表 6：スレッド4（「レス」等の用語）の主成分分析結果

第1主成分	固有値 20.7 累積寄与率 14.5%	主成分の意味 記事の長さ
負荷の高いキーワード	人, 思, 意味, 的, 性, 言葉, 用, 一, 特, 見, 場合, 文, 話, 以, 論	
負荷の低いキーワード	時間, 車, 週, 誌, 刻, 流, 外, 記, 悪, 相手, 誤, 足, 様, 取, 度, 個	
第2主成分	固有値 5.1 累積寄与率 18.0%	主成分の意味 主観：客観
負荷の高いキーワード	自分, 理, 文, 事, 間, 点, 相手, 関係, 全, 最, 意見, 指摘, 論, 持, 力	
負荷の低いキーワード	広, 指, 状, 一般, 用法, 現, 用, 定着, 意味, 少, 混乱, 誤用, 言葉, 区, 正, 例	
第3主成分	固有値 4.9 累積寄与率 21.4%	主成分の意味 知性：感性
負荷の高いキーワード	指摘, 関係, 論, 議論, 例, 学, 呼, 性, 日, 点, 特, 然, 関, 要, 場合	
負荷の低いキーワード	聞, 最, 自分, 知, 識, 問, 人, 大, 見, 違, 意, 持, 結, 事, 当, 言	
第4主成分	固有値 4.0 累積寄与率 24.2%	主成分の意味 現在：歴史
負荷の高いキーワード	区, 別, 呼, 時, 間, 言語, 時間, 刻, 会, 車, 単, 然, 表, 気, 分	
負荷の低いキーワード	無, 読, 明, 誤用, 一般, 行, 坂口, 得, 足, 記事, 指摘, 主張, 用, 取, 説, 広	

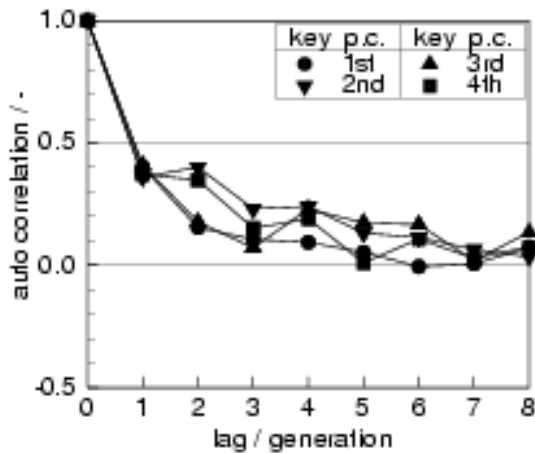


図 6: スレッド4（「レス」等の用語）の自己相関係数

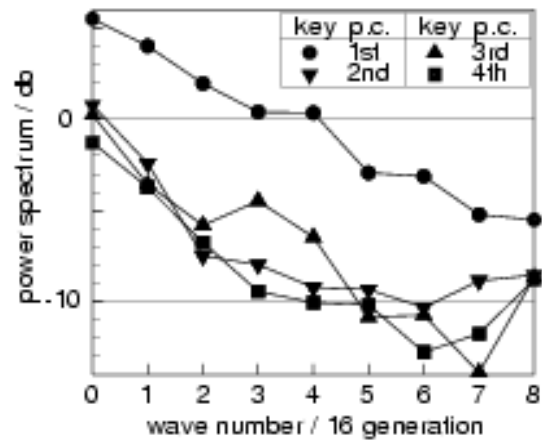


図 7: スレッド4（「レス」等の用語）のパワースペクトル

見感性的とも思われる「意地悪」というキーワードが低い負荷を示しているが、このスレッドの中でこの用語は感性的な反発に対する冷静な分析を行う文脈で用いられている。

相関係数を図 8 に、パワースペクトルを図 9 に示す。第 3 主成分に 2 世代を周期とする顕著な振動が認められる。なお、第 2 主成分にも振動が存在する可能性があるが、ノイズとの識別は難しい。

第3主成分スコアが高い値を示す記事の多くは同一の人物から投稿されており、第 3 主成分スコアの上位 13 位までの記事は、全て同一人物が投稿している。この投稿者は、スレッド 6 を構成する 456 記事の約 20% に相当する 94 記事を投稿しており、その投稿姿勢はスレッド全体の特性に影響を与えていると考えられる。その投稿内容は end 行に批判的な立場であるが、主張の

異なる投稿者に対する執拗な批判がなされ、これが第 3 主成分の振動を拡大している可能性がある。

第 3 主成分スコアの低い記事は、「意地悪」という感性的用語の出現頻度も高いが、その内容は end 行を意地悪とする主張に対する反論であり、普遍性を重視し判断に慎重さを求める傾向が認められる。

このスレッドにおける議論を読むと、二つの価値観の対立が明瞭に認められる。end 行を記事に含めることを問題なしとする主張はインターネットの規約 (RFC) を基準に判断し、普遍性を重視する。一方、記事に end 行を含めるべきでないとする主張は、多くの人が使う MSIN で特定の記事に問題が起こるならそれは記事が悪いと判断し、普遍性よりも利便性を重視する。これら二つの相反する主張は、いずれも相手の理解を得ることができず議論は平行線をたどった。

表7: スレッド6 (end 行の理由) の主成分分析結果

第1主成分	固有値 17.6 累積寄与率 15.5%	主成分の意味 記事の長さ
負荷の高いキーワード	思, 記事, 人, 事, 考, 言, 私, 意見, 出, 目, 違, 的, 一, 当, 他	
負荷の低いキーワード	日本語, 嫌, 動, 馬鹿, 初, 文字化, 田, 時田, 話, 使, 化, 句, 力, 方法, 為, 字	
第2主成分	固有値 4.7 累積寄与率 19.6%	主成分の意味 指摘方法: 感性的反応
負荷の高いキーワード	字, 引用, 文, 文字化, 部, 変, 示, 単, 以, 部分, 定, 作, 上, 性, 指摘	
負荷の低いキーワード	方, 事, 議論, 悪, 然, 良, 意地悪, 自, 言, 人, 嫌, 入, 間, 他, 思, 問題	
第3主成分	固有値 3.6 累積寄与率 22.8%	主成分の意味 意味: 普遍性
負荷の高いキーワード	知, 出, 来, 行, 方法, 為, 作, 見, 質問, 実, 別, 力, 返, 話, 理由	
負荷の低いキーワード	良, 意地悪, 主, 少, 目, 議論, 文字化, 意見, 手, 的, 数, 思, 考, 多, 無, 間	
第4主成分	固有値 3.6 累積寄与率 25.9%	主成分の意味 感性: 知性
負荷の高いキーワード	自分, 内, 容, 考, 勝手, 気, 日, 言, 時田, 大, 悪, 通, 論, 他, 田	
負荷の低いキーワード	的, 定, 意地悪, 理由, 事, 自, 良, 知, 少, 多, 者, 主, 行, 場合, 意味, 使	

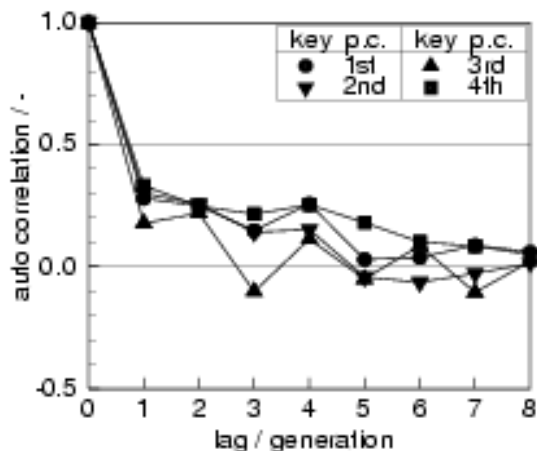


図8: スレッド6 (end 行の理由) の自己相関係数

本手法で振動が検出された主成分は、「MSIN ユーザに読めないend行を含む記事でMSINの問題点を指摘するのは不当である」とする、実用的効果を重視した強い主張と、普遍性を重視してend行を不当とする判断に慎重さを求める主張とを両極端とする成分であり、スレッドの内容から読み取られる対立点にほぼ対応する。但し、対立する一方の側は一人の投稿者であり、その用語使用癖が分析結果にかなりの影響を及ぼしている可能性もある。

6. 考察

以上のように、4つのスレッドに我々の手法を適用した結果、一部の自己相関係数にラグ2を周期とする明らかな振動が認められた。また、これに対応するパワースペクトルにも、2世代を周期とする点で上昇が認めら

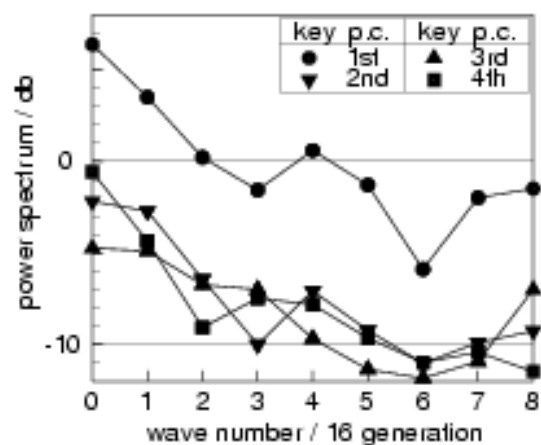


図9: スレッド6 (end 行の理由) のパワースペクトル

れ、2世代を周期とする主成分スコアの振動が存在すると判断された。相関係数0.2に対する95%信頼区間は、スレッドを構成する記事数が400の場合、0.10 ~ 0.29と見積もられ(国沢(1966)[19])、振動の検出は統計的にも有意である。

ラグ2を周期とする自己相関係数の振動と、2世代を周期とするパワースペクトルとは、同一の現象を意味する。これは、一連の議論において、ある基準で分類したとき、異なる範疇に属する用語の出現頻度が高いメッセージが交互に現れていることを意味する。各々の主成分に対応する用語の分類基準は、主成分負荷の高い用語と低い用語を参照すること、および、主成分スコアの高い記事と低い記事内容の読解によって把握された。

我々が分析したような大きなスレッドは、ほぼ例外な

く何らかの対立する主張を含んでいる。しかし、特性指標が振動するケースもあれば、振動しないケースもあった。議論における対立点が主成分スコアの振動に現れなかったケースとして、スレッド1(署名の引用)及びスレッド4(「レス」等の用語)が挙げられる。この2つのスレッドに共通する特徴として、議論の中心には対立する主張が存在するものの、双方の主張が他方の主張を尊重する形で議論が行われた点が挙げられる。このような形で議論がなされると、双方のメッセージには類似したキーワードが現れる傾向が高いため、主成分スコアの振動が観察されなかったものと思われる。

一方、議論における対立点と主成分スコアの振動成分が一致したケースとして、スレッド3(初心者への対応)及びスレッド6(end行の理由)が挙げられるが、この2つのスレッドのいずれにおいても自己の論理を繰り返し主張し、対立する主張に対する理解が不足している傾向が認められる。

前者の対立においては、確かに対立する主張が存在するが、噛み合った議論が行われて、必ずしも参加者全ての納得が得られないまでも、参加者の合意ないし共通認識の獲得に至っており、ネットニュースというコミュニケーション手段自体は充分有効に利用されていると考えられる。一方、後者の対立においては、対立する主張が存在するだけでなく相互の意思の伝達がなされておらず、コミュニケーション自体が有効に機能していないと言える。後者では、対立する二つの集団が、コミュニケーションによって解消しえない異なった文化的背景を持っていることを見いだせたとも言える。

本稿では、ネットニュース上の対立関係を定量的に抽出する手法として、キーワード出現度数の主成分分析に基づく手法を提案した。ネットニュース上の大きなスレッドにこの手法を適用して分析した結果、一部のスレッドの特定の主成分に振動が認められた。

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、コミュニケーションにおける相互理解不能という、より深刻な対立関係である。大勢の人々が、かなりの回数の意見交換を重ねても、互いに分かり合うことができないという事態は、文化的背景やそれによってたつ集団の価値観や常識とでもいうべきものが大きく異なっているということを意味する。意見交換を重ねることにより解消できる対立とは質的な違いのある対立を定量的に抽出することにより、二つの相反する

文化を抽出したと言えるかも知れないが、その抽出した二つの文化がどのようなものであるのかについての解釈は、今後の課題としたい。

本稿では分析手法を中心に扱い、実際のネットニュースでなされている議論の分析は一部のスレッドについてのみ行ったが、今後は範囲を拡大して分析を行い、電子的なコミュニケーションを介して形成された新しい社会の特性を明らかにしたいと考える。

謝辞

本研究に際しては、東京大学先端経済工学研究センター児玉文雄教授のご指導をいただきました。ここに感謝の意を表します。

注および参考文献

- [1] McLuhan, M., "Understanding Media: The Extensions of Man", McGraw-Hill, New York, 1964. (栗原裕, 河本伸聖訳「メディア論—人間の拡張の諸相」, みすず書房, 1987)
- [2] Sproull, L.; Kiesler, S., "Connections: new ways of working in the networked organization", The MIT Press, 1992. (加藤丈夫訳, 「コネクションズ—電子ネットワークで変わる社会」, アスキー出版局, 1993)
- [3] Johnston, K.; Johal, P., "Internet as a virtual cultural region", *Internet Research: Electric Networking Applications and Policy*, Vol.9, No.3, pp. 178-186, 1999.
- [4] 川上善郎; 川浦康至; 池田謙一; 古川良治, 「電子ネットワークの社会心理—コンピュータ・コミュニケーションへのパスポート」, 誠信書房, 1993.
- [5] 瀬尾雄三, 「ネットニュースにおける投稿行動の解析」, 筑波大学経営・政策科学研究科修士論文, 1996.
- [6] 村越広享; 落水浩一郎, 「電子メールを利用した協同作業における会話のコヒーレンスと話題の完結度の関係」, コンピュータソフトウェア, Vol.15, No.3, pp. 50-53, 1998.
- [7] 村越広享; 島津明; 落水浩一郎, 「電子メールを利用したコミュニケーションにおける討議の流れの自動抽出について」, 情報処理学会研究報告, Vol.FI51-2, pp. 9-16, 1998.
- [8] 井佐原均; 小作浩美; 内元清貴, 「討論型ニュースグループを対象とする知的ニュースリーダーの開発」, 情報処理学会研究報告, Vol.NL119-3, pp. 13-18, 1997.

- [9] 内元清貴；小作浩美；井佐原均，「対話型ネットニュースグループにおける話題転換記事の推定」，言語処理学会第3回年次大会発表論文集，pp. 377-380, 1997.
- [10] 内元清貴；小作浩美；井佐原均，「キーワードによるネットニュース記事群の構造化」，言語処理学会第4回年次大会発表論文集，pp. 544-547, 1998.
- [11] 斎藤典明；水澤純一；山本平一；山口英，「話題の自動抽出による電子メールの情報組織化手法」，情報処理学会論文誌，Vol.39, No.10, pp. 2907-2913, 1998.
- [12] 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)
- [13] 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.
- [14] 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回年次大会発表論文集，pp. 209-212, 1999.
- [15] Horton, M.; Adams, R., "Standard for Interchange of Usenet Messages", Network Working Group, RFC-1036, 1987.
- [16] 日野幹雄，「スペクトル解析」，朝倉書店，pp.25-27, 1977.
- [17] <ftp://mitsuko.jaist.ac.jp/pub/news-archive>.
- [18] nao1@urban.ne.jp, Re: 初心者です，63puo9\$5lq@nntp.urban.or.jp, 1997.
- [19] 国沢清典，「確率統計演習2—統計」，培風館，1966.