

# 掲示板における炎上分析のためのノイズ除去手法

## Denoising techniques for flaming analysis of BBS

前川 和輝†  
Kazuki Maekawa

松本 和幸†  
Kazuyuki Matsumoto

吉田 稔†  
Minoru Yoshida

北 研二†  
Kenji Kita

### 1. はじめに

近年、インターネット上のソーシャルネットワークサイトや電子掲示板等で、多数の人々が互いの意見を気軽に交換できる環境が整備されている。このような場において交わされる議論には、人々の関心を引くものからあまり関心を引かないものまで幅広く存在する。2ちゃんねる掲示板などの匿名掲示板において自由に繰り広げられる議論により、軽率に発言された失言、誹謗中傷や風説の流布などが、個人や企業活動などに影響を与えることもある。大きな問題に発展する前に、問題の存在を認識し、適切な対策を実施することが求められる。しかし、インターネット上には様々なユーザによる多数の掲示板サイトが存在しており、日々多くの議論がなされている。このため、膨大な掲示板データのなかから、すべての議論の内容を、人手により詳細に確認することは、コスト面において現実的ではない。膨大な議論の中から問題として発展しそうな重要な議論を自動的に抽出できれば、人的、時間的コストの削減に有効と考えられる。

問題として発展しそうな重要な議論の一つとしてネット炎上の事例が挙げられる。一般に、ネット炎上とは、ブログやTwitter等のSNS上において発言された軽犯罪自慢や誹謗中傷などが原因で、その内容が瞬く間にWeb上で拡散され、2ちゃんねる等の匿名掲示板でスレッドが乱立し、その結果、個人では收拾がつかなくなってしまう状態のことを指す。このように、ネット炎上には匿名電子掲示板が大きく関わっており、そこでの当該話題の盛り上がりが多いユーザの注目を集める要因となる。こうした炎上の原因となる盛り上がりの発端となる発言を検出するなどの分析をおこなうためには、話題の遷移における変化を検出する必要がある。

しかし、炎上状態においては、次々に書き込まれてスレッドが消費されていき、新たなスレッドが立てられるといった特性上、話題の遷移を迫るために、膨大な量の書き込みの確認が必要となる。スレッドのなかには、話題の本筋とは関係の無い書き込みも含まれるため、確認作業が困難となる。内容に乏しいと考えられる書き込みとして、つぎのようなものが考えられる。

- ・ 視覚的効果を狙ったアスキーアートの書き込み
- ・ スレッドの話題とは無関係の書き込み
- ・ 一人の参加者によるスレッドの大半を占める同じ内容による書き込み

本研究では、上記のような、話題の流れを把握する妨げとなる書き込みをノイズとして定義し、これらをフィルタリングする手法を提案する。この手法により、ユーザによる書き込み内容の把握を支援できる。

### 2. 関連研究

文書の要約の研究は従来から盛んにおこなわれている。Kleinbergの提案するburst検出手法に基づいた手法では、ブログや電子掲示板における書き込みを、時間情報を含む文書列であるdocument streamとして扱う。この手法では、あるdocumentの出現間隔が短くなっている箇所(burst)に、注目すべきイベントが存在すると判定し、検出をおこなう。この手法では、イベント発生と無関係にdocument数が変動するようなdocument streamをうまく扱うことができないため、藤木ら[1]は、時間区間を極限まで短くすることにより拡張したトピックワードの抽出手法の有効性を評価している。これらの手法は、話題の盛り上がった瞬間を検出するものであるが、本研究では炎上の原因分析が目的であるため、盛り上がりの前後の状態も調べる必要がある。

### 3. 提案手法

本研究では、話題の移り変わりを数値化するため、スレッド単位での重要語抽出をおこなう。文字重複度、文字連続度に関しては、あるしきい値より大きい場合にフィルタリングをおこない、単語エントロピーに関してはあるしきい値より小さい場合にフィルタリングをおこなう。書き込みごとのフィルタリングの後、単語の品詞を条件として絞りこみ、単語頻度にもとづきtf-idfによる重み付けをおこなう。ノイズフィルタリングから単語の重み付けまでの処理の流れを図1に示し、図に従って各処理について説明する。

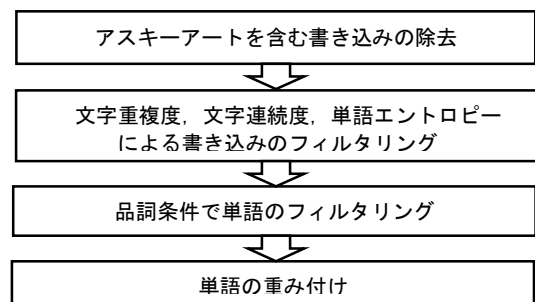


図1 フィルタリング処理工程

#### 3.1 アスキーアートのフィルタリング

アスキーアートによる書き込みは、スレッドの内容に大きく影響を与えるような重要な情報はほとんど含まれないと考えた。また、アスキーアートには半角スペースと全角スペースの連続が含まれるものが多いことから、これら2つの文字の組み合わせが一度でも含まれる書き込みを、アスキーアートが含まれていると判定し、除去する。

#### 3.2 文字重複度によるフィルタリング

アスキーアートには同じ文字が多数使用される。また、アスキーアートのほかにも、同じ文字が多数出現する書き込みには無意味なものが多いと判断した。

†徳島大学, Tokushima University

文字列  $s$  の文字重複度を表す  $\text{dupscore}(s)$  は、式(1)により計算される。文字列  $s$  は、 $n$  個の文字から構成され、 $\text{uniq}(s)$  は、文字列  $s$  中の文字の種類数を示す。

$$\text{dupscore}(s) = \frac{|s|}{|\text{uniq}(s)|} \quad (1)$$

$$s = \{c_i | 1 \leq i \leq n\}$$

$\text{uniq}(s) \cdots$  文字列  $s$  中の文字の種類数

### 3.3 文字連続度によるフィルタリング

文字重複度と同様、同じ文字が連続する割合が大きい書き込みには、無意味なものが多いと判断した。式(2)に示す  $\text{seqscore}(s)$  は文字列  $s$  の文字連続度を示す。

$$\text{seqscore}(s) = \frac{|s|}{\sum_{i=1}^n \text{dif}(c_i, c_{i+1})} \quad (2)$$

$\text{dif}(c_i, c_j) \cdots c_i$  と  $c_{i+1}$  が異なる場合に値 1 をとる

$c_{n+1} \cdots s$  中のどの文字とも異なる終端文字

### 3.4 単語エントロピーによるフィルタリング

現在 2 ちゃんねるのスレッドが盛り上がっているかどうかの指標として、「勢い」というものがある。これは、ひとつの掲示板において書き込み速度(消費速度)を計算することで得られる。しかし、単にスレッドを消費させることが目的と考えられる、内容に乏しい無意味な書き込みまで盛り上がり度を上昇させることがある。掲示板における無意味な書き込みとしてコピー&ペーストが多く見られる。それらの書き込みは書き込みの情報量(エントロピー)が小さいのではないかと考えた [2]。

### 3.5 単語の重み付け

スレッドの書き込み内容を特徴付ける単語について、TF-IDF 法を用いた重み付けをおこなう。

## 4. 検証

ある大学生が Twitter 上で不適切な発言をしたために炎上した話題についてのスレッド(2011年)の勢いとエントロピーの関係を表す図 2 を用いて検証する。このなかで、特に特徴のある①~④に着目する。①はもっとも勢いのある箇所、②はエントロピーが比較的小さい箇所、③はもっともエントロピーの小さい箇所、④は勢いの盛り返しをもっとも大きい箇所である。また、スレッド間のフィルタリングから得たキーワードの上位 50 件の一致の推移を、キーワード一致率として図 3 に示す。一致率は、提案手法を適用し、連続する 2 つのスレッドからそれぞれ抽出した上位  $N$  件のキーワード集合を  $K_i, K_{i+1}$  と表すと、式(3)により表すことができる。

$$\text{一致率}(i, i+1) = \frac{|K_i \cap K_{i+1}|}{N} \quad (3)$$

図 3 から、図 2 において、勢いのある 1 スレッドから 22 スレッドまでにおいては 52 スレッドから 76 スレッド間までと比べて一致率の割合の変動が小さい。特に、10 スレッドから 11 スレッドにおいて話題の転換が起こったと予想される。後半の 52 スレッドから 76 スレッド間では話題の転換が起こっては同じ話題で盛り上がっていることが図 2 から読み取れる。

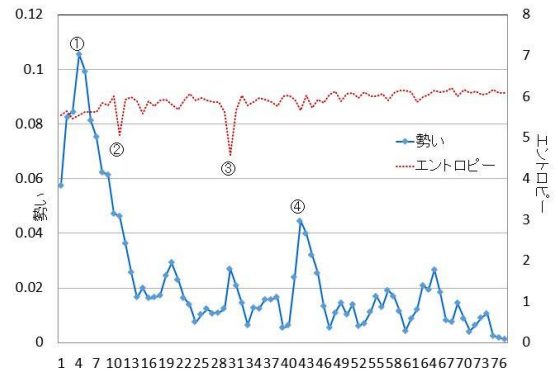


図 2 炎上スレッドの勢いとエントロピー

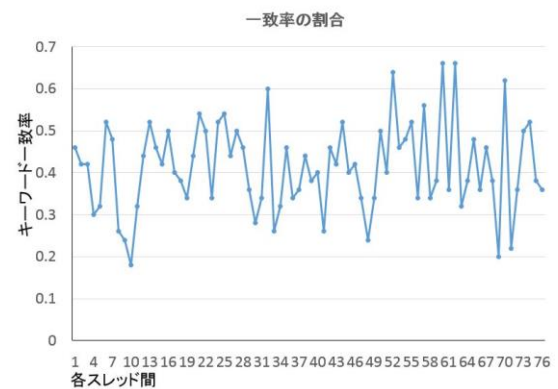


図 3 スレッド間のキーワードの一致率

## 5. おわりに

本稿では、掲示板における炎上分析のためのノイズ除去手法として、単語エントロピー、アスキーアート、文字重複度、文字連続度による書き込み単位での除去および、品詞等による単語単位での除去を提案した。長期間継続してシリーズ化しているスレッドの場合、日にちをまたいだ発言者の ID が変動し、発言者の特定が難しいことから、本稿では同一参加者による書き込みを考慮したフィルタリングはおこなわなかった。しかし、短時間では参加者 ID が固定されることを利用したフィルタリング手法を提案したいと考えている。今後は、得られた単語がスレッドのキーワードとなりうるのかを評価する方法を検討し、炎上スレッドだけでなく他のスレッドにも応用できるように改良したい。

### 謝辞

本研究の一部は、科学研究費補助金(基盤研究(C)15K00425、若手研究(B)15K16077)の補助を受けて行った。

### 参考文献

- [1] 藤木稔明, 南野朋之, 鈴木泰裕, 奥村学: "document stream における burst の発見", 社団法人 情報処理学会 研究報告, NL-160, 2004.
- [2] 前川和輝, 松本和幸, 吉田稔, 北研二: "書き込みのエントロピーに着目した速報スレッドにおける盛り上がり要因の分析", 言語処理学会第 21 回年次大会発表論文集, pp.620-623, 2015.