

オンラインディスカッションの健全度定量化手法の提案

A Method for Quantifying Soundness of Online Discussion

武吉 朋也† 帆足 啓一郎† 松本 一則† 滝嶋 康弘†
Tomoya Takeyoshi Keiichiro Hoashi Kazunori Matsumoto Yasuhiro Takishima

1. まえがき

SNS (Social Networking Service) 等のオンラインコミュニケーションサービスの普及により、近年では一般ユーザでもオンラインで容易にディスカッションを行える。このような状況下では、健全に盛り上がるディスカッション以外に、正常に進行されない荒れたディスカッションとなる場合がある。荒れたディスカッションが多くなると、ユーザがサービスを利用しなくなる一因となるため、サービスの運用者はその場を健全に保つための行動が必要となる。しかしながら、多数のディスカッションについて、健全に盛り上がっているのか、荒れているのかを判断するには、現状目視で確認する必要がある、多大なコストを要する。

ディスカッションの荒れを測定する既存手法として、文献[1]が挙げられる。この手法は健全なディスカッションに出現する単語のリスト、および荒れているディスカッションに出現する単語のリストを用い、入力されたディスカッションでの両リストに含まれる単語の出現頻度を基に、荒れている度合いを定量化する。しかしながら、リストに含まれる代表的な単語の出現頻度に依存するため、リスト以外の単語を用いて荒れる場合には、ディスカッションの健全度を正しく定量化できない。また、筆者らは、発言間の時間間隔、および各発言で新規に出現する単語数に着目し、ディスカッションの健全性の判断を支援する手法を提案した[2]。さらに、評価実験の結果、健全性を正しく判断するには、各発言の有益性やユーザの発言行動の特徴といった観点での指標が不足しているという課題を明らかにした。

本稿では上記の課題を解決するため、各ユーザや各発言、発言間で定義される指標を追加し、さらに既存のディスカッションデータを学習用データとし、健全度を定量化する手法を提案する。提案手法の応用により、運用者が荒れたディスカッションを抽出するためのコストの削減、更にはユーザが参加するディスカッションを選ぶ際の判断基準の提供や、参考情報として健全なディスカッションを提示することが可能となる。以降では、提案手法について述べ、さらに既存のディスカッションデータを用いた評価実験の結果と考察について述べる。

2. 健全度定量化手法

本章では、まず提案手法の概要について述べ、次に、ディスカッションの特徴を表す指標について述べる。

2.1. 手法の概要

提案手法は、ディスカッションの健全度を定量化するモデルを学習する処理と、学習したモデル(以降、健全度定量化モデル)を用いて新たに入力されたディスカッションの健全度を定量化する処理からなる。以降、各処理について説明する。

健全度定量化モデルの学習

ディスカッションデータから算出され、その特徴を表す複数の指標を説明変数、人手により付与されたディスカッションの健全度を目的変数とし、多変量解析を適用することで健全度定量化モデルを学習する。本稿では多変量解析として線形回帰を適用し、式(1)で表される健全度定量化モデルを学習する。

$$y = c + \sum_{i=1}^k a_i x_i \quad (1)$$

なお、式(1)の y は健全度、 c は線形回帰の切片、 a_i は i 番目の指標に対する偏回帰係数、 x_i は i 番目の指標値、 k は用いている指標の総数を表す。

健全度定量化

健全度定量化の対象となるディスカッションデータが入力されると、データから指標値 x_1, x_2, \dots, x_k を算出する。次に、健全度定量化モデル(式(1))へ各指標値を入力し、健全度 y を定量化する。

2.2. 指標の概要

本稿では、ディスカッション内の各発言について、その発言内容と発言ユーザ名、および発言時刻が得られる場合に算出可能な指標のみを用いる。これにより、汎用的な健全度定量化を実現する。また、各発言に定義する指標については、特定の文字(疑問符等)の出現頻度や形態素解析結果を用いて、特定の単語の出現に依存しない指標とする。以降、提案手法で用いる指標の一部について列挙する。なお、(i)と(ii)は筆者らが以前[2]で提案した指標である。

- (i) 発言間の時間間隔。
- (ii) 各発言に含まれる単語のうち、新規出現の単語数。
- (iii) 発言しているユーザの総数。
- (iv) 各ユーザの発言数。
- (v) 各ユーザについて、そのユーザ名が発言内容に含まれた頻度。
- (vi) 各発言に含まれる文字数。
- (vii) 各発言に含まれる単語数。
- (viii) 各発言に含まれる文数。
- (ix) 各発言に含まれる命令形の動詞数。
- (x) 各発言に含まれる疑問符「？」の数。
- (xi) 同一ユーザによる連続発言数。
- (xii) 2名のユーザによる交互発言数。
- (xiii) 連続する発言間の共起単語数。

ただし、全てのユーザ、全ての発言や発言間で合計値、平均値、中央値、最小値、最大値が算出可能な指標については、これらを全て指標値とする。例えば、(iv)各ユーザの発言数については、合計値がディスカッション内の総発言数となり、平均値は各ユーザあたりの平均発言数を表し、これらの指標値を健全度定量化モデルの説明変数に加える。提案手法では、全部で248個の説明変数を用いる。

† (株) KDDI 研究所, KDDI R&D Laboratories Inc

3. 評価実験

本章では、健全度定量化手法の有効性を確認するための評価実験を行い、その結果について述べる。まず、実験に用いたデータと正解の定義等、実験条件について述べる。次に、実験結果と考察を述べる。

3.1. 実験条件

本稿では、Wikipedia¹のノートページ上で行われているディスカッションを評価実験に用いる。ここで、ディスカッションの単位は、ノートページ中の「節(section)」で区切られるテキスト範囲とし、発言の単位については発言ユーザ名と発言時刻で区切られるテキスト範囲とした。なお、総発言数が15以上30以下であるディスカッションを実験用データとした。

実験時の正解ラベルとして、各ディスカッションに下記3種類のラベルを、被験者5名により付与した。なお、各ラベルにはそれぞれ括弧内の数値を健全度として対応させた。

(ラベル1)「ほぼ全体が健全」(健全度 90)

(ラベル2)「一部が荒れ」(健全度 50)

(ラベル3)「ほぼ全体が荒れ」(健全度 10)

ただし、「荒れ」とは、言い争いや平行線の発言、他のユーザを攻撃する発言等が含まれ、正常に進行していない場合として、被験者に教示した。2037個のディスカッションデータに対してラベル付与を行った後、対応する健全度について5名の平均値を算出し、各ディスカッションの健全度とした。最終的に得られたディスカッションの健全度を、健全度定量化モデルの学習時には目的変数、検証時には正解値とした。

なお、本稿の実験においては、交差検証として Leave One Out²を行った。学習用データから得られた線形回帰モデル(式(1))に検証用データの指標値を入力し、予測値としての健全度を算出した。また、正解の判定基準として、健全度が80以上であるデータは「健全」、20未満であるデータは「荒れ」、これら以外は「一部荒れ」を正解カテゴリとし、各データに対する予測値についても同様の基準で予測カテゴリとした。

3.2. 結果と考察

提案手法により各ディスカッションの健全度を予測し、正解カテゴリとの一致数を算出した。まず、全体での一致率は72%であった。さらに、正解が「健全」であるデータを「荒れ」と誤って予測すること、および正解が「荒れ」であるデータを「健全」と誤って予測することは無かった。これらに加え、提案手法では容易に算出可能な指標のみを用いていることを鑑みると、正答率は十分高いと考えられる。また、カテゴリ別の結果を表1に示す。なお、表の適合率は各カテゴリでの予測数に対する一致数の割合を表し、再現率は各カテゴリでの正解数に対する一致数の割合を表す。適合率では、「健全」カテゴリの適合率が76%と高く、

「荒れ」カテゴリの適合率が低い。「健全」カテゴリの適合率が高いことから、参考情報としてユーザに健全なディスカッションを提示するといった利用シーンを想定する場合、提案手法は有効に作用すると考えられる。また、再現率では、特に「荒れ」カテゴリの再現率が17%と低い結果になった。荒れたディスカッションをサービス運用者に提示する利用シーンを想定する場合、高い再現率が求められるため、提案手法に更なる改善が必要である。

そこで、「荒れ」カテゴリの再現率が低い原因を調査するため、正解カテゴリが「荒れ」であるデータのうち、健全度の予測値が最も高かったデータを確認した。このデータでは、参加ユーザ数が16人と多く、かつほとんどの発言が他者の発言を考慮することなく、自身の意見を述べるだけであり、ユーザ同士のやり取りが行われていなかった。一般に、ディスカッションでは他者の意見を理解するための質問や質問に対する回答、他者の意見を受けての自身の意見を発言内容に含む。提案手法では、2名のユーザが交互に発言している回数や、連続する発言間での共起単語数を指標として用いているが、複数ユーザ間のやり取りの内容について、ユーザ間で定義される指標を用いていない。そのため、ユーザ同士のやり取りが行われていないディスカッションについては、健全度を正しく定量化できていないと考えられる。

表1 提案手法の正答率

	健全	一部荒れ	荒れ
適合率	76% (370/489)	71% (1068/1508)	55% (22/40)
再現率	53% (370/700)	89% (1068/1205)	17% (22/132)

4. まとめ

本稿では、オンラインで行われているディスカッションの健全度を測るため、ディスカッションの特徴を表す複数の指標を定義し、健全度を定量化する手法を提案した。提案手法は、各発言について発言内容と発言ユーザ名、および発言時刻が取得できれば、適用可能な汎用的な手法である。既存のディスカッションデータに提案手法を適用した結果、全体での正答率は72%であるが、荒れたディスカッションに対する適合率、再現率は共に低いため、今後改善が必要である。

今後は、荒れたディスカッションに対する抽出精度向上のため、ユーザ間での発言内容の類似度といった、特定のユーザの発言内容に限定し、ユーザ間で定義される指標の追加を検討する。また、特定の単語の出現頻度に基づく手法[1]について、本稿と同様の評価実験を実施し、提案手法との比較を行う。

参考文献

- [1] Yu Ichifuji, Susumu Konno and Hideaki Sone, "A Method to Monitor a BBS Using Feature Extraction of Text Data," Proc of 3rd International Conference on Human.Society@Internet, LNCS 3597, pp.349-352, (2005).
- [2] 武吉朋也, 帆足啓一郎, 松本一則, 滝嶋康弘, "オンラインディスカッションの効率性定量化手法の提案," 第72回情報処理学会全国大会, (2010).

¹ Wikipedia 日本語版のデータベースダンプのうち、2009年12月20日版を下記サイトよりダウンロードし、使用した。 <http://download.wikimedia.org/jawiki/>

² n 個のデータのうち n-1 個のデータを学習用データとし、残りの1個を検証用データとする交差検証法。