

電子掲示板のスレッド内の話題に注目した参照関係を可視化する研究

中村健二[†] 田中成典[‡] 高橋亨輔[†] 田中裕一[†] 今城彰子[‡]

関西大学大学院総合情報学研究科[†] 関西大学総合情報学部[‡]

1. はじめに

Web におけるコミュニケーション手段として電子掲示板が提供されている。電子掲示板では、消費者による意見交換や評価が行われ、膨大な情報が集まる場となっている。しかし、1つのスレッド内では複数の話題が議論されるため、スレッド内の各意見と話題との関係性を把握し難しく、可読性が低いという問題[1][2]がある。可読性の向上に関する既存研究には、電子会議のログから話題を表わす語を抽出し、話題語の位置づけを重要度のグラフ上に可視化する研究[3]や、会議議事録から会議中の意見の推移を可視化する研究[4][5]がある。しかし、前述の既存研究[1]-[5]では、話題ごとに意見が整理されている電子会議や会議議事録を対象としており、電子掲示板のような話題と意見が混在している対象への適用は難しいという問題がある。そこで、本研究では、電子掲示板を対象に、各意見の関係性を解析し話題ごとに分類して可視化することで、電子掲示板の可読性を向上する手法を提案する。

2. 研究の概要

本提案手法では、電子掲示板の1スレッド分のHTMLデータを入力し、電子掲示板特有の参照ルールと話題語の伝搬過程に着目した解析を行い出力する。出力は、話題と意見の関係図を木構造に構造化した図として作成する。システムの概要を図1に示す。本システムは、1) HTML解析機能、2) 話題ごとの意見の分類機能と3) 意見の参照関係可視化機能により構成される。

2.1 HTML解析機能

本機能では、電子掲示板のHTMLを解析し掲

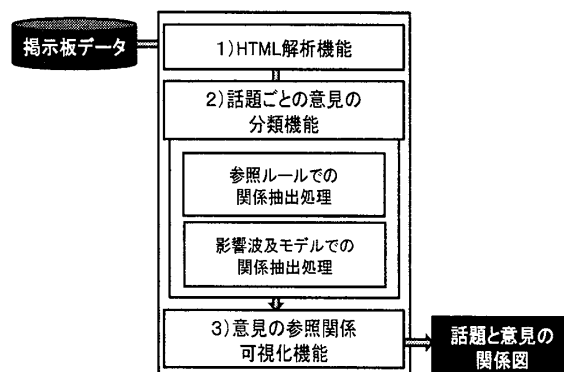


図1 システムの概要

示板内の意見を意見ごとに投稿者名、意見内容、投稿日時を抽出する。

2.2 意見の参照関係抽出機能

本機能では、参照ルールでの関係抽出処理と、影響普及モデルでの関係抽出を行い、話題ごとに意見を分類する。

2.2.1 参照ルールでの関係抽出処理

本処理では電子掲示板が持つ特有の参照ルールを用いて意見間の参照関係を抽出し、関係性のある意見同士を結び付ける。電子掲示板特有の参照ルールには、特定の話題に対する意見を示すため、話題を提示している意見の本文に「>>」という記号を付与して引用するルールや、投稿者の名前を「○○さん」のように記載するルールがある。

2.2.2 影響普及モデルでの関係抽出処理

本処理では、影響普及モデルを用いて意見が属する話題を判定する。影響普及モデルとは、話題語の伝搬性をモデル化したものである。影響普及モデルを用いることで電子掲示板の意見中で新しく現れた語を話題を表わす話題語と判断し、その伝搬性を用いて参照関係を抽出する。

2.3 意見の参照関係可視化機能

本機能では、話題ごとの意見の分類機能の結果を基に、電子掲示板の意見を話題ごとに構造化し、木構造で関係図を可視化する。

Research for Visualization of Relationship Between Topic Opinions Within Bulletin Board System

[†] Kenji Nakamura, Kyosuke Takahashi

Graduate School of Informatics, Kansai University, 2-1-1 Ryouzenji-cho Takatsuki-shi, Osaka 569-1095, Japan

[‡] Shigenori Tanaka, Yuichi Tanaka, Akiko Imagi

Faculty of Informatics, Kansai University, 2-1-1 Ryouzenji-cho Takatsuki-shi, Osaka 569-1095, Japan

3. システムの実証実験と考察

本システムの実行結果を図2に示す。本提案手法の有用性を検証するために、手作業で作成した正解データと本システムにより可視化した結果について比較実験を行う。

3.1 実証実験

本実験では、価格.comのロコミ掲示板の中でも、コメント数が30件以上のスレッド8件を対象とした。本実験で出力した関係図について、人手で作成した関係図と比較し、意見同士が正しく結びつけられているかどうかの評価を行った。

3.2 結果と考察

図2の結果よりを確認すると掲示板内の意見が構造化されて表現されていることが分かる。また、本提案手法で抽出したスレッド毎の正解意見数と適合率を表1に示す。表1の結果より、意見数が少ないスレッドでは、高い適合率を示していることが分かる。これは、意見が少数であるほど意見が集約しているためである。また、スレッドCは意見数が多いにもかかわらず、高い適合率を得ることができた。これは、掲示板特有のルールが多く使用されていたためである。

また、全体を通して、掲示板特有のルールで結び付いている意見は正確に取得できており、参照関係を特定するための有効な手段であるといえる。それに対して、意見数が多いスレッドA, B, DとEでは、適合率が低い結果となった。これは、影響普及モデルが原因であると考えられる。影響普及モデルは参照ルールを持たない意見を木構造に組み込む点では効果を発揮したが、単語で重み付けを行っているため、意見が持つ単語数に影響を受けやすいという欠点がある。そのため、誤判定となった意見は、影響普及モデルにより単語を多く含む意見に結びつけられ、誤判定となったと考えられる。

4. おわりに

本研究では、意見が属する話題を考慮して、電子掲示板の可視化する手法を提案した。本提案により、電子掲示板内の意見を話題ごとに構造化して可視化することができた。これにより、掲示板の可読性が低いという問題を解決した。しかし、本研究では意見の内容を考慮せず可視化しているため、重要度の低い意見と高い意見が同列に表示されている。そのため、十分な支援ができていないとはいえない。今後は、意見の内容を考慮することで、電子掲示板の可読性を向上する方法について研究する予定である。

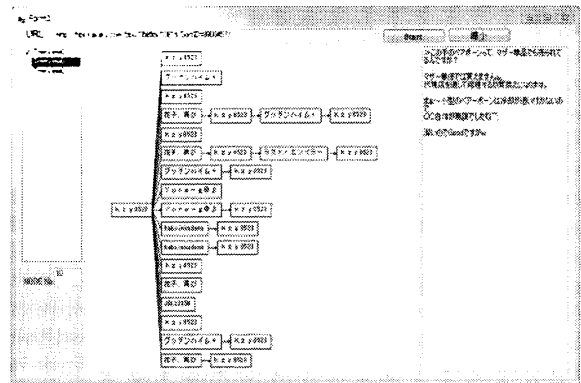


図2 実行結果

表1 正解ノード数と適合率

	意見数	正解意見数	適合率
スレッドA	52	25	0.48
スレッドB	85	35	0.41
スレッドC	48	45	0.94
スレッドD	78	31	0.39
スレッドE	49	18	0.38
スレッドF	8	8	1.00
スレッドG	13	13	1.00
スレッドH	22	18	0.81
平均	49	25	0.51

参考文献

- [1] 望月俊男, 久松真一, 八重樫文, 永田智子, 藤谷哲, 中原淳, 西森年寿, 鈴木真理子, 加藤浩: 電子会議室における議論内容とプロセスを可視化するソフトウェアの開発と評価, 日本教育工学会論文誌, 日本教育工学会, Vol.29, No.1, pp.23-33, 2005.7.
- [2] Spiegel, D.: A Visualization of the Conversational Dynamic within IRC, Master's Thesis, Massachusetts Institute of Technology, 2001.10.
- [3] 藤井敦: OpinionReader 意思決定支援を目的とした主観情報の集約・可視化, 電子情報通信学会論文誌, 電子情報通信学会, Vol.J91-D, No.2, pp.459-470, 2008.2.
- [4] 柴田裕介, 山口和紀: 議論の論理構造に着目した議論分析手法の提案, 人文科学とコンピューター研究会研究報告, 情報処理学会, Vol.2007, No.9, pp.41-48, 2007.1.
- [5] Smith, M.: Tools for Navigation Large Social CyberSpace, Communication of the ACM, Association for Computing Machinery, Vol.45, pp.51-55, 2002.4