

森のメタフォを用いた掲示板における 重要人物・返信数・重要単語の可視化

小栗 崇志[†], 津田 耕平^{††}, ラック ターウオンマツト^{††}

[†] 立命館大学理工学部

^{††} 立命館大学大学院理工学研究科

本稿では、近年多くの人々が利用している掲示板について、発言者の中で特に重要な人物、返信数、重要単語などを可視化することでよりわかりやすくする手法について述べる。可視化については、発言者同士のお互いの返信数から、ほかの発言者との距離を求めて座標データを作る。その座標データの上に我々が提案する森のメタフォに基づいて木を配置する。提案のメタフォでは、人物を木で、その重要度、返信数、重要単語をそれぞれ幹の大きさ、葉の数、果実の数で表す。

Visualization of important person, the number of replies, and important words in bulletin board services using forest metaphor

Takashi OGURI[†] Kouhei TSUDA^{††} Ruck THAWONMAS^{††}

[†] College of Science and Engineering, Ritsumeikan University

^{††} Graduate School of Science and Engineering, Ritsumeikan University

In this paper, we propose a technique for bulletin board service visualization using forest metaphor. In particular, important persons, the number of replies, and important words are visualized. In this metaphor, a tree represents a user, the number of leaf corresponds to the number of replies, and a fruit indicates an important word.

1 はじめに

現在、掲示板は多くの人たちに利用されている。しかし、掲示板への書き込みは、すぐさま書き込むことができ便利であるが、書き込みを行うユーザも多く、書き込みの量が増えればとても複雑なものになってしまう。また、どの人物が中心で掲示板での議論が行われているのかわからず、だれに対して返信をすればいいのかもわかりづらい。このような場合に、議論から外れているユーザに返信してしまい、返信が返ってこなかったり、返ってきても期待した返信ではなかったりする場合がある。このような問題は、誰を中心として議論が行われているか、中心となっている議論は何か、ということがわかればある程度解決できることだと

思われる。

そのときに有用となってくるのが可視化であり、実際に掲示板を花に例えて可視化している研究がある [1]。他にも、掲示板の時間軸に注目して可視化している研究 [2] や、人間の身体の特徴に例えている研究もある [3][4]。また、「可視化」は電子メールやニュースグループといったものも対象とし、さまざまなことが可視化されている [5][6]。そういった「可視化」という手法に注目し、本研究では掲示板を対象とすることにした。

そこで本論文では、ある掲示板中のデータをもとに、木のメタフォをつくり、掲示板をひとつの森として表す手法について提案する。以下では、2章で可視化の際に扱う掲示板のデータについて、3章で掲示板でのユーザ同士の距離、4章で森のメ

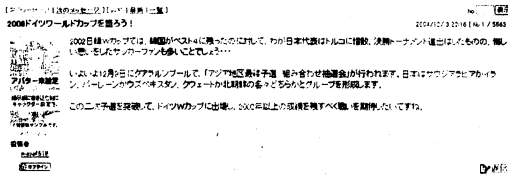


Fig. 1 掲示板

タフォを用いた可視化について述べ、5章でまとめて述べる。

2 掲示板データ

2.1 データ抽出

ここでは、今回扱ったデータについて説明する。まず、実際の掲示板 [7] からそれぞれのユーザの投稿数、返信数、重要単語といったデータを抽出する。実際に用いた掲示板を図 1 示す。それぞれのデータに対する説明は以下の通りである。

投稿数とは、そのユーザが掲示板に書き込んだ総数である。返信数とは、そのユーザが書いた書き込みに対して、他のユーザから返信された総数である。返信に関しては、単に返信された総数だけでなく、それぞれのユーザから合計何回返信されたかというユーザ別返信数というデータも抽出する。また、投稿数、返信数という二つのデータから、1件あたりの返信数というデータも求める。この1件あたりのデータ数というのは、返信の総数を投稿数で割ったもので、このデータも後述する4章の可視化部分で使用する。重要単語に関しては2.2で述べる。

2.2 重要単語

重要単語は、その掲示板で書かれている単語の中で、中心となっている議論を表している単語であることが多い。そして、その重要単語とは、全コメントに対して「茶釜」[8]という形態素解析システムを用いて名詞を抜き出し、その結果をキーグラフ [9] というツールにかけて求めたキーワードである。図 2 で示しているのが、キーグラフにかけた結果である。

3 ユーザ同士の距離

ここでいうデータの距離とは、掲示板中で、ユーザ同士がお互いにどれくらい返信し合っているか

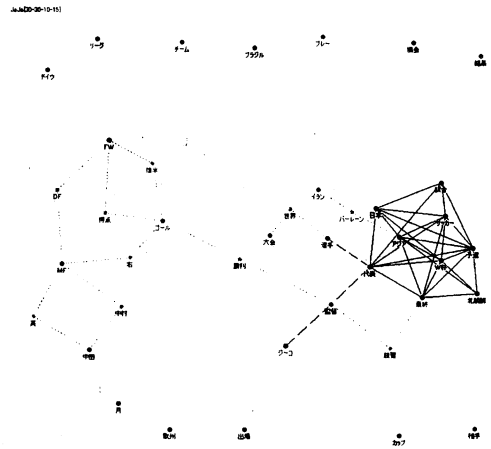


Fig. 2 キーグラフ

で定義している。それぞれのユーザごとに誰に何回返信しているかというユーザ別返信数から、それぞれのユーザごとに距離を求める。この距離の計算には、ユーザ同士のユーザ別返信数の少ないものを使用する。そして、この距離は近ければ近いほど返信しあっているということになり、以下はユーザ A とユーザ B との距離の式である。

$$d_{AB} = \begin{cases} \frac{\alpha}{\min(r_{AB}, r_{BA})} & \min(r_{AB}, r_{BA}) > 0 \\ \alpha + 1 & \min(r_{AB}, r_{BA}) = 0 \end{cases} \quad (1)$$

ここで α は正の整数、 r_{AB} 及び r_{BA} はそれぞれユーザ A がユーザ B に返信した数及びユーザ B がユーザ A に返信した数を表す。どちらか一方の返信しかない場合にプラス1したのは、一番遠い距離 α のときよりも遠くに設定するためである。例としては、ユーザ A がユーザ B に対して 40 件返信し、反対にユーザ B がユーザ A に対して 20 件返信したのであれば、 $\min(r_{AB}, r_{BA})$ は 20 となり、式の上段が適用される。これに対して、ユーザ A はユーザ B に 5 件返信したが、ユーザ B はユーザ A に対して返信を 1 件もしなかった場合は、 $\min(r_{AB}, r_{BA})$ は 0 であり、式の下段が適用されることになる。

3.1 座標マップの作成

可視化の際には、先述した距離に基づいて木を配置する。木の配置のために座標マップが必要であり、

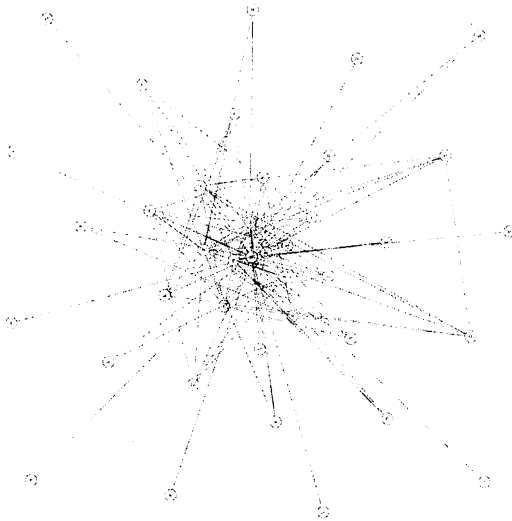


Fig. 3 座標マップ

この座標マップの作成にあたっては Graphviz[10] というツールを使用して作成した。その図については図3で示している。

4 可視化

4.1 割り当て

ここでは可視化の割り当てについて述べる。割り当てとは、掲示板のデータを可視化する際、木、葉、幹などの何に置き換えるかということである。可視化中で使っているものは、葉、木の太さ、木の高さ、果実、木同士の距離である。

まず、葉の数はあるユーザが掲示板中に書き込んだ数（投稿数）を用い、掲示板に書き込んでいるユーザほど葉が多くなる。次に、木の太さ、高さはあるユーザの1件あたりの返信数である。1件あたりの返信数は、返信数を投稿数で割ったものであり、この1件あたりの返信数が多いほど木が高く、太くなる。果実は掲示板中の重要単語を表したもので、果実が多いほど、多くの種類の重要単語を発言しているということである。今回、果実の数はひとつの木について最大で15となっている。木同士の距離は、3章で述べたように、ユーザ同士の距離であり、距離が近ければ近いほどそのユーザ同士は返信し合っているということである。

上記の割り当てを適用した可視化のサンプルは



Fig. 4 木のサンプル

図4に示す。左の木については、右の木に比べ葉と果実が多く、木のサイズは小さいため、書き込みを多くしており、重要単語も書き込んだ記事の中に多く含まれているが、書き込みひとつあたりの返信数は少ないということになる。右の木については、左の木に比べ葉と果実が少なく、木のサイズは大きいいため、書き込みをしている数と重要単語は少ないが、書き込みひとつあたりの返信数は多いということになる。

4.2 可視化プログラム

可視化する際のプログラムは JavaApplet で作成した。Java Applet List for Forest[11] というサイトで紹介されている、木をランダムで生成するプログラムをベースに、先述した、投稿数や返信数などのデータをファイルで読み込み、木を生成する場所、葉や果実の数等、こちらの目的にあったプログラムに改良した。

4.3 森

図5は可視化の結果を示す。今回は $\alpha=10$ で計算している。実際はカラーの図であるが、形式上、白黒で出力している。ちなみに、カラーの場合、葉が緑、木が茶色、果実が赤という配色になっている。

4.4 考察

実際の掲示板からデータを抽出し、可視化するというところまで完成することができた。それぞれのデータが何を表しているかは、4.1をみていただきたいが、図5を見ることで、中心にある木が、葉も多く、果実も多いため、中心となっている人

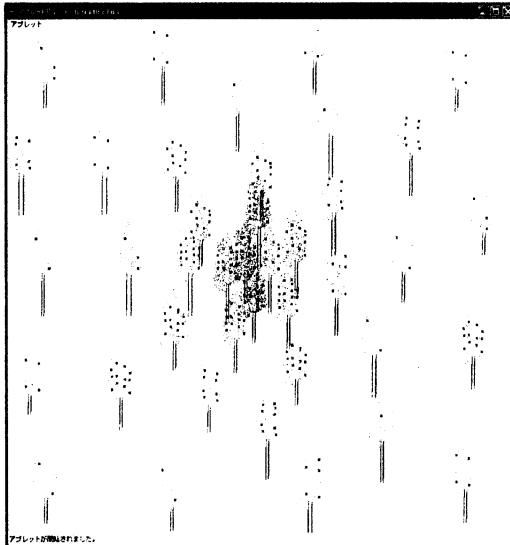


Fig. 5 可視化

物達だとわかる。また、それぞれの距離がとても近いことから、中心にいる人物達は、お互いに返信し合っている関係だということもわかる。反対に、画面の端のほうにある木は葉が少なく、果実も少ないため、書き込みがあまり行われておらず、話題となっている議論について述べられていないことがわかる。また、中心となる木から距離が離れているため、その掲示板の中心人物ではないこともわかる。

5 おわりに

本稿では、掲示板を対象として取り上げ、掲示板の可視化について提案を行った。具体的には、掲示板から投稿数や返信数といったデータを抽出し、実際に可視化するプログラムを JavaApplet で作成した。可視化では木や葉というメタフォを使ったが、感覚的に理解しやすいものにできたと思われる。

参考文献

- 1) Rebecca Xiong., Judith Donath.: PeopleGarden: Creating Data Portraits for Users. ACM Symposium on User Interface Software and Technology 1999: 37-44.
- 2) Rehman Mohamed., John Ferguson., David

Elsweiler., Alasdair Mac Cormick., John Wilson., George Weir.: BulB - VISUALISING BULLETIN BOARD ACTIVITY. ICEIS (4) 2004: 31-35.

- 3) Ethan Perry., Judith Donath.: Anthropomorphic Visualization: A New Approach For Depicting Participants in Online Spaces. CHI Extended Abstracts 2004: 1115-1118.
- 4) Ethan Lewis Perry.: Anthropomorphic Visualization: Depicting Participants in Online Spaces Using the Human Form. Media Arts and Science 2004.
- 5) Fernanda B. Viegas., danah boyd., David H. Nguyen., Jeffrey Potter., Judith Donath.: Digital Artifacts for Remembering and Storytelling: PostHistory and Social Network Fragments. HICSS 2004.
- 6) Fernanda B. Viegas., Marc Smith.: Newsgroup Crowds and AuthorLines: Visualizing the Activity of Individuals in Conversational Cyberspaces. HICSS 2004.
- 7) Yahoo!掲示板,
<http://messages.yahoo.co.jp/bbs?.mm=SP&action=m&board=1834688&tid=2006a5ia5a4a5da5oa1bca5ka5ia5aba5ca5wa4r8la4ma4a6a1aa&sid=1834688&mid=1&type=date&first=1>
- 8) 形態素解析システム茶筌,
<http://chasen.naist.jp/hiki/ChaSen/?FrontPage>
- 9) キーグラフ, <http://www.chokkan.org/xoops/modules/tinycontent/index.php?id=5>
- 10) Graphviz, <http://www.graphviz.org/>
- 11) JavaApplet, http://www.gifu-keizai.ac.jp/~sugihara/set/CLlist_f.htm