

# 電子掲示板からの興味ある会話の抽出支援<sup>\*1</sup>

岡村 剛, 角 康之, 西田 豊明

京都大学

## 1. はじめに

Web 上には多数の電子掲示板があり、そこでの話題は様々である。大規模な掲示板では、ほかの媒体では得られないような「くちコミ」の情報も多い。しかし、発言の内容は非常に雑多であるため、その中から自分の興味に即したものを探すのは大きな負担となる。

本論文では、電子掲示板での発言の重要度や会話の意味構造を可視化し、そこからユーザの興味のある会話を抽出する手法を提案する。

## 2. 掲示板での会話の意味構造の可視化

電子掲示板の形式として一般的なものに、ツリー形式がある。話題を提示する発言や、各発言に対する返信を書き込むことができる。この形式では、興味のある発言のツリーをたどることで、その話題についての発言が得られる。しかし、返信先が1つに限られるため、同じ話題の発言が複数のツリーに分散することもあり、大規模な掲示板を見る際には負担となっている。

議論支援システム AIDE[1]では、発言同士の関係を、テキストオブジェクトの空間配置によって表現している。空間の自動構成には、双対尺度法[2]と呼ばれる統計的手法を採用している。これは、複数の数量化属性で構成されたオブジェクト集合を与えられたときに、オブジェクト同士の属性共有性と属性同士の共起関係を顕在化するように空間を構成する手法である。各発言に対してキーワードベクトルを与え、双対尺度法を用いることで、発言とそれらを構成するキーワードの関連性を、同一の距離空間上に配置することが可能となる。これによって、複数の発言を受けた発言でも、それらのつながりを表現することができる。ここで、発言とキーワードのクラスタを、発言のツリーに対応するものとして見ることができる。

## 3. 発言、キーワードの重要度

文書内での語の重要度を定量化する手法として、tfidf法があげられるが、これを電子掲示板の発言に適用した場合、単に珍しい語が重要とみなされたり、一般的な語であるが話題の中心となっている語が重要でないといみなされてしまう恐れがある。

影響の普及モデル[3]では、返信関係からなるツリーを伝播していくキーワードに注目し、ある返信 B 中のキーワードのうち、返信先 A に含まれるものの割合を A からの影響量としている。また、B への返信 C があった場合、A から C への影響量は、C のキーワードのうち、A・B ともに含まれるものの割合に、A から B への影響量をかけたものとして定義する。ある発言からその下のツリーに伝播した影響量の総和をとることで発言が媒介した影響量が得られる。また、影響量は各語が均等に媒介していると仮定すると、各発言間での1語あたりの影響量が求められる。各語についてその総和をとることで、各語が媒介した影響量が得られる。

本論文では、発言やキーワードの重要度を定量化する方法として、この手法によって得られる影響量を採用した。

## 4. 会話の意味構造、発言・キーワードの重要度の可視化

価格.com(<http://www.kakaku.com>)の「ニンテンドーDS」に関する使用レポートの掲示板について、会話の意味構造、発言・キーワードの重要度の可視化を試みた。ここでは、球にキーワードまたは発言の番号をラベルとして付与したものを配置している。また、この掲示板では、レポートが肯定的・否定的であるかどうかタグ付けしてあり、それぞれ緑、赤と色づけした。キーワードは青としている。

まず、影響量が上位の語からなるキーワードベクトルを発言ごとに求め、双対尺度法を適用する。それにより得られた空間のうち、相関係数の上位2軸を用いて発言とキーワードの配置を決定した。そして、影響度によって球の大きさを決定した。

<sup>1</sup> Supporting extraction of a dialogue from a bulletin board system according to user's interest  
Go Okamura, Yasuyuki Sumi, Toyooki Nishida  
Kyoto University

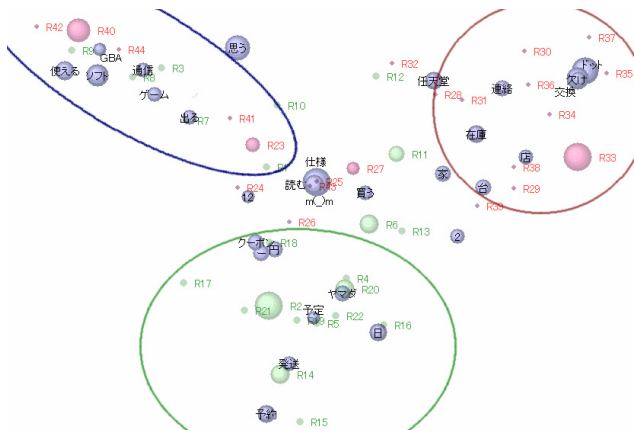


図 1：価格.com の発言・キーワードの意味構造と重要度を可視化した例

## 5. 可視化された空間の妥当性

図 1 を見ると、肯定的な発言・否定的な発言・両方の発言がそれぞれ集まっているクラスタがあることがわかる。それぞれを①肯定クラスタ（緑の楕円）、②否定クラスタ（赤の楕円）、③対立クラスタ（青の楕円）と呼ぶことにする。

クラスタ内の影響度の大きいキーワードは、①予約・クーポン・電器店名、②ドット欠け・交換、③ソフト・使える・GBA・通信などである。

実際に発言の内容を確かめると、①は、電器店での購入報告・在庫状況・Web 通販での予約と発送日など、購入そのものについての話題、②は、液晶画面のドット欠けについての話題、③は、面白いソフトの話題と、GBA のゲームも遊べるが、通信対戦機能に対応していないという話題であり、キーワードの組み合わせから想起されるものと一致していた。「ドット欠け」と「交換」は空間内で非常に近い位置にあったが、「液晶にドット欠けがあったので販売店で交換してもらった」という発言がいくつかあり、実際に関連の深いキーワードが近くに配置されていたことが分かる。

クラスタ内の影響度の大きい発言だけを見ると、実際にその話題の起点となっており、また、どのクラスタにも含まれていなかった発言には返信が少なかったり、あっても話題が変わっていたりと、比較的盛り上がっていない話題であった。

以上のことから、肯定的なレポート・否定的なレポートについて、主な話題はどのようなものか、どの発言が重要であったかということが、可視化された空間に実際に反映されていたと言える。

## 6. 発言の抽出

くちコミによる商品の評価は、購入の意志決定

においては有用である。複数の商品から 1 つを選びたいという場合も多い。その際にはそれぞれの長所・短所を比較して決めることになる。知人からの口コミは、その人の興味・視点などがある程度わかっているので、より共感できる人の意見を重視するといったことができるが、電子掲示板での口コミにはそのような背景情報はない。そこで、視点ごとに発言を整理することを考える。

まず、各商品の掲示板について同じキーワードベクトルを用いて空間を構成する。それぞれの対立クラスタからは、商品の長所についての発言とそれに対する反論が抽出できる。ある商品の肯定クラスタと他の商品の否定クラスタが重なっていても、ある商品が良いという根拠となる発言が抽出できる。また、複数の商品の肯定クラスタが重なっていても、この視点ではそれらの商品どれもが良いという根拠となる発言が抽出できる。

## 7. 会話の抽出

6. で抽出した発言は、各商品の長所・短所をあらわすものである。パソコンのスペックのように客観的な情報であれば、それらを並べてユーザーに提示することも考えられる。しかし、口コミの情報は主観的なものが多く、ただ並べて眺めるだけではわかりにくい。また、抽出した発言は商品を肯定する発言と反論など、会話的な性質をもっている。そこで、各商品を勧める仮想的な人物どうしの会話として表現できれば、それぞれがどのような視点から商品进行评估しているかを明らかにすることができ、理解しやすいのではないだろうか。

各発言は 6. のように整理することで、どの人物に割り振るか、どの発言の応答として利用できるかといったことがわかる。

今後の課題は、4. の可視化された空間が人間の感覚とできるだけ一致するようにすることと、抽出した発言をつなげ、自然な会話を生成する手法を考案することである。

## 参考文献

- [1] 角康之：議論の意味構造の可視化。可視化情報学会誌, Vol. 19, No. 72, pp. 7-12, 1999.
- [2] 西里静彦：質的データの数量化—双対尺度法とその応用—。朝倉出版, 1982.
- [3] 松村真宏, 大澤幸生, 石塚満：テキストによるコミュニケーションにおける影響の普及モデル。人工知能学会論文誌, Vol. 17, No. 3, pp. 259-267, 2002.