

CGMのテキスト解析とマーケティングへの活用

○佐藤邦弘
(日経リサーチ)

鈴木督久
(日経リサーチ)

1. はじめに

近年、ブログ、SNSに代表されるCGMの登場により、Web上には一般的な消費者が作成したテキスト型のデータが急増している。武田(2006)はCGMの中でも特にブログの特徴について6点にまとめているが、本発表ではその中でも特に次の3点に注目をした。

- ① 同一(あるいは少数)の人物によって書かれている
- ② 時系列表示されている
- ③ 他の記事やページへのリンクを多くもつ

このようにブログの情報は、IDで紐付けられ(①)、時系列データをもち(②)、ID間のコミュニケーション情報が付与された(③)、高度に階層化されたテキストデータといえる。本セッションでは、ブログデータをもちいて、このような構造をもつブログデータを定量と定性の両面から活用する方法を提案する。

2. ブログデータの収集

ブログデータの収集までの流れを図1にまとめる。ブログの投稿後に生じるコミュニケーションを追跡するため、トラックバックが存在する場合には、新しいURL集合としてセットし、最大3次先までのブログ記事の追跡を行なった。

3. ブログ上でのクチコミの定量分析

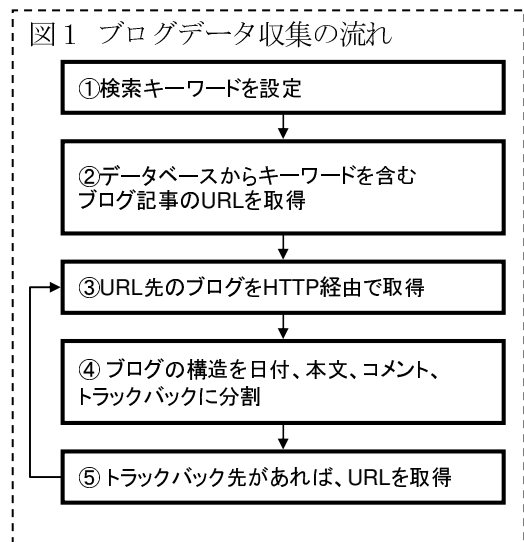
オンラインコミュニティで消費者によって作成された評判情報はクチコミを介して他の消費者へと影響を与える(小川(2003))。購買への影響が口頭で交わされるクチコミは測定するのが極めて難しい一方で、ブログのようなアプリケーションを介して交わされるクチコミは解析可能なデジタルデータとしてリアルタイムに観察することができる。そこで、ブログ上に顕在化したクチコミに限定し、クチコミの定量化指標としてコメントやトラックバックなどのコミュニケーション機能に注目した3つの指標(①露出量、②伝播量、③密集度)を作成し、ブログデータを解析した。

まず、①の露出量はキーワードを含むブログの記事数を表す。この作成されたブログ記事が他のブロガーやROMの目に触れるため、この量を露出量と呼ぶことにする。

次に、②の伝播量は、ブログ記事にブログ記事についてのコメントの数を足し合わせたものである。ブロガーが作成したブログ記事に対して関心が寄せられた場合、寄せられるコメントの数は多く、その逆は少ないと考えられる。ブログ記事としてブロガーが最初に提供した話題が周囲にどれだけ広がったかを表すことから、この量を話題の拡がりを表す量として伝播量と呼ぶことにする。

最後に、③の密集度は、ブロガー間のコミュニケーションネットワークの密度を示す。特定のキーワードに対し、ブログ上のコミュニティが反応した場合、相互にコメントやトラックバックのやりとりが行なわれる。このため、コメントやトラックバックのやりとりによって発生するコミュニケーションの数が、相対的にそのコミュニケーションへの参加人数と比べ小さくなる。このことを用いてコミュニケーション

図1 ブログデータ収集の流れ



の密度を以下のように定義する。

$$\text{密集度} = (\text{記事を書いたブロガーの人数} + \text{ブロガー間の関係総数}) / \text{参加ブロガー総数} \quad - 1$$

結果

キーワード「Lumix」で収集したブログデータを上記の3指標を用いて分析した時系列の結果を次に示す。

図3-1 2007年1月 ブロガーネットワーク (密集度=0.57)

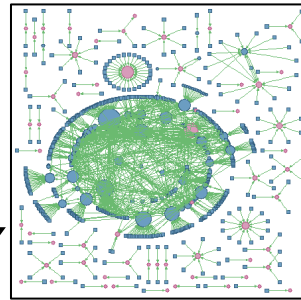


図3-2 2007年3月 ブロガーネットワーク (密集度=0.05)

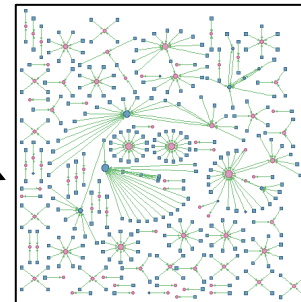


図2 キーワード「Lumix」の時の3指標の時系列推移

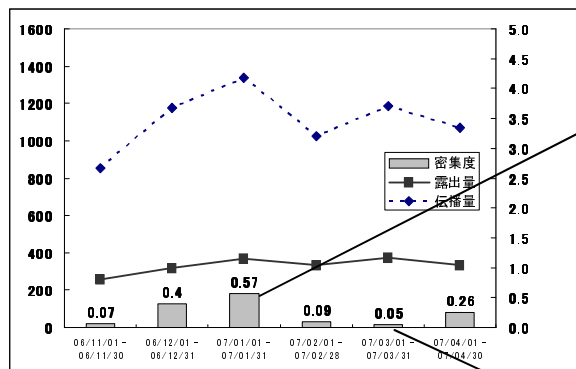


図2は3指標の時系列推移を表している。2007年1月のデータは、露出量に比べて伝播量が大きく上昇しており、このことはブログ上でのコメントのやりとりが活発になっていることを示している。また、図3-1と図3-2は密集度が高いときと低いときの違いをブロガー間のコミュニケーションネットワークで比較したものである。露出量、伝播量との間に大きな差がなくても、密集度が高い場合には、非常に密なコミュニケーションが発生し、密集度が低い場合は、コミュニケーションが疎となっていることが分かる。

4. まとめ

CGM上で発生するクチコミを解析する手段として、ブログのコミュニケーション機能に着目し、①初期反応数：露出量、②会話の広がり：伝播量、③密度：密集度という3つの視点から多面的にクチコミの測定を試みた。その結果、コミュニティの反応や会話の盛り上がり度合いなど、ブログ記事の投稿数のみからは表れないクチコミの一面を指標から得ることができた。

参考文献

- 武田英明(2006), 『Weblog 研究の現状』, 人工知能学会, 第7回セマンティックウェブとオントロジー研究会資料.
- 小川美香子, 佐々木裕一, 津田博史, 吉松徹郎, 國領二郎(2003)「黙って読んでいる人達 (ROM) の情報伝播行動とその購買への影響」『マーケティングジャーナル』, Vol. 22, No. 4, pp. 39-51.