

ブログをマーケティングデータとして活用するために

佐藤邦弘

株式会社日経リサーチ マーケティングリサーチ本部 Web調査企画グループ チームリーダー

1.国内ブログの現状

今から4年前の2005年、「ブログ」はユーキャンの新語・流行語大賞のトップ10に選ばれるなど大きなブームになっていた。ブログはその後順調に成長を続ける。2008年時点で、月に1回以上更新されるアクティブブログは約300万、ひと月に投稿されるブログ記事に至っては4000万～5000万件という膨大な量に至っている（総務省情報通信政策研究所「ブログの実態に関する調査研究」より）。ブログはもはや一過性のブームでは無く、ひとつのメディアとして定着したと言えるだろう。

ブログの登場は、それまで情報発信力の弱かった普通の消費者に強力な情報発信ツールをもたらした。ブログを手に入れた消費者は日常のちょっとした話題を日々発信している。このように、ブログの登場は“消費者発のメディア”と“新しいデータソース”の2つの意味から、消費者を巻き込んだ新しいマーケティングが必要とされる時代を予感させるものだった。

2.ブログデータの活用を阻む障害

このように日々、大量の情報が発信・公開されているブログだが、この大量の“消費者の“声””をマーケティングのデータとして活用する取り組みはまだ始まったばかりである。ブログのデータは高いポテンシャルを秘めているものの、障害も多い。

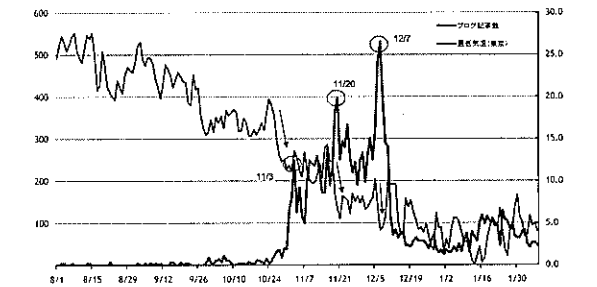
最初に障害となったのは、アフィリエイトやSEOでの効果を狙ったスパムと呼ばれるブログがあまりに多いことだった。ただ、スパムブログについては、当初から問題が認識されていたこともあり、自動判定技術の進歩などにより状況は改善しつつある。障害の次の焦点は、ブログデータそのものの特性へと移ってきている。

ブログデータの特性としては、そのほとんどをテキストデータが占めていることが挙げられるだろう。そのデータもブロガーによって自由に書かれた文章の為、アンケートの自由回答データやカスタマーセンターに集まったテキストデータ等と比べると、知りたい情報と掛け離れた話題も多く、正確なカテゴリが困難な不安定なデータとなっている。その一方で、安定して分析できる定量データは少ない。オンラインでブログから取得できる定量データとしては、①投稿日、②記事数、③サイト数、④テキストの場所（タイトル/記事本文/コメント文）、⑤コミュニケーションデータ（誰が誰にコメントを返したのか）といったところがほとんど全てだろう。性別・年齢や購買経験の有無といった足りない情報は、ある程度のエラーは黙認し自動判定技術を使うか、コストを掛けて人力で判定することになる。このことから、ブログデータをマーケティングデータとして有効に活用していくためには、この膨大なテキストデータをいかにうまく分析できるかが鍵となる。以下で、具体的な分析事例を使って説明したい。

3.“ババシャツ”と呼ばれていた「ユニクロ・ヒートテック」

2008年の冬に爆発的なヒットとなった「ユニクロ・ヒートテック」。そのブログでの書き込みはどのようなものだったのだろう。図表1は、ヒートテックについて書かれたブログの記事数と東京の最低気温の推移を追ったグラフである。ヒートテックについての書き込みは、11月3日、11月20日、12月7日に大きく増えており、いずれもその少し前に最低気温が大幅に下がり寒くなっていることが分かる。このことはブログの記事数の変動が消費者の購買行動などと密接な関係があったことを示唆していると言えるだろう。次に、図表2で変動した要因を単語の

図表1 「ヒートテック」を含むブログ記事数と最低気温（東京）の時系列推移



書かれ方の変化によって試みる。まず、ブームが始まる前の10月1日の時点では、「ヒートテック」は「ユニクロのババシャツ」という説明付きでブロガーに書かれていることがわかる。その後、11月になり気温が下がって寒くなると、購入者を中心に「暖かい」といった書き込みが増え、12月には「噂」「人

図表2 「ヒートテック」と関連して書かれていた単語の推移

