

CGM コンテンツの分析は集合痴に要注意

三菱総合研究所
飯尾 淳

集合知や CGM(Consumer Generated Media) は、IT の活用領域としてまだまだ発展する余地が大きい分野である。今後も様々なサービスが開発されていくだろう。本コラム¹でも、これまで何度か集合知や CGM に関するテーマを取り上げている。集合知や CGM という話題、いまだホットな領域であり、目を離せない分野である。

しかし一方で、集合知や CGM のプラットフォームたり得る掲示板やコメント欄、twitter などのオンライン・コミュニケーションツールは、炎上²したり、田代砲³のような悪戯が蔓延したりといった危険も孕んでいる。そのような状況はもはや「集合知」ではなく「集合痴」という名がふさわしい。集合知は新しい知識発見の方法として非常に魅力的だが、そこに集合痴が混じると道を誤るリスクがある。

集合痴になってしまったケースか否かの見極めは、生のデータを人間が少しでも読めば、それほど難しくないだろう。しかし機械的な判断は、さほど簡単な話でもない。大量の CGM コンテンツをデータマイニングやテキストマイニングのツールで機械的に処理しようとする場合には、集合痴になってしまったケースを自動で排除する仕掛けを用意しておかなければ、誤った結論を導いてしまいかねない。そのためには、集合痴状態か否かを自動で判断する基準として使うための、何らかの指標を用意しておく必要がある。

集合知から集合痴に至る道

具体的な例を示そう。本稿では、書評の分析を題材にして、集合知と集合痴の区別を考えてみたい。

本稿で以下に示すグラフは、大手オンライン書店サイトに寄せられたユーザーレビュー、つまりユーザーによる書評を分析したものである。およそ 500 件以上のレビューが寄せられている書籍をいくつかピックアップし、その傾向を調べてみた。以下に示すグラフの縦軸は各レビューの文字数、横軸は、そのコメントの有用度である。なお縦軸は対数目盛りで刻

まれている点に注意されたい。このグラフに、書評と共に示されている星の数による評価で分類した各レビューを、プロットした。

なお有用度は「そのコメントが参考になったか否か」の投票で定められる。ユーザーはレビューコメントを読んで、「参考になった」もしくは「参考にならなかった」との意見を投稿できる。そこで、「参考になった」という投票を投票の総数で割った値として、有用度を定義する。有用度は 0.0 から 1.0 の値をとる。その値が高いほど、そのレビューの価値が高い、そのレビューに共感したユーザーが多い、あるいは、そのレビューに対する支持者が多い、ということの意味する。有用度を、そのレビューに対する支持率と読み替えても構わないだろう。

では 6 作品の分析結果を順にみていこう。順番は、投稿されているレビュー数の少ないほうから並べている。

まともな集合知のケース

ケース 1: 最初に示すグラフは、いま非常に人気の高いミステリー作家の手による直木賞受賞作に寄せられたレビューの分布である (図 1)。この作家の

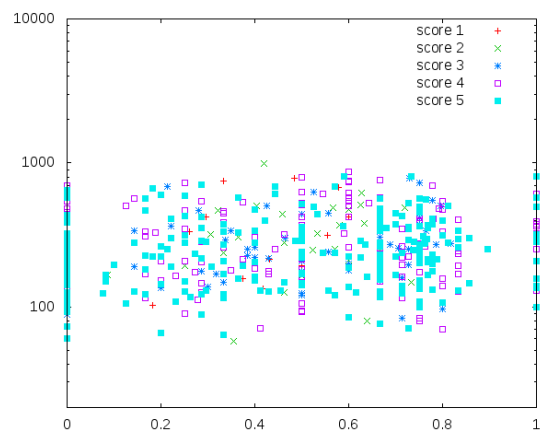


図 1: オンライン書店サイトに投稿された書評の分析 (ケース 1: まともな集合知)。書評と共に示された星の数 (score) 別に、その書評の支持率 (横軸) と文字数 (縦軸、対数目盛り) との関係がプロットされている。

¹本稿は、週刊 Take IT Easy (<http://easy.mri.co.jp/>) 上に昨年発表されていたものを、配付資料用にセミナー主催者が改編したものである。

²ブログのコメント欄や掲示板が燃え上がること

³TIME 誌の Person of the year に端を発する現象

作品に対する書評は星5つと高い評価が多く、星4つ、星3つが徐々に減っていく逆三角形の分布をみせているという特徴を持つ。星2つや星1つという評価は少ない。映像化された作品も多く、高い人気を誇るベストセラー作家だけに、愛読者による真面目なレビューが多く投稿されていることの証であろう。それだけにレビューコメント自体にも真面目な評価が加えられている。グラフをみると、星の多寡に関わらず、コメントの評価が一様に分散している傾向が分かる。

ケース2およびケース3: 次は、グラフを2つ示そう。1つは昨年ベストセラーとなったビジネス書ふうの小説(図2)、もう1つは昨年夏に映画が公開され、驚きのラストシーンが待っていたミステリー小説だ(図3)。両者に共通して、先のグラフよりは

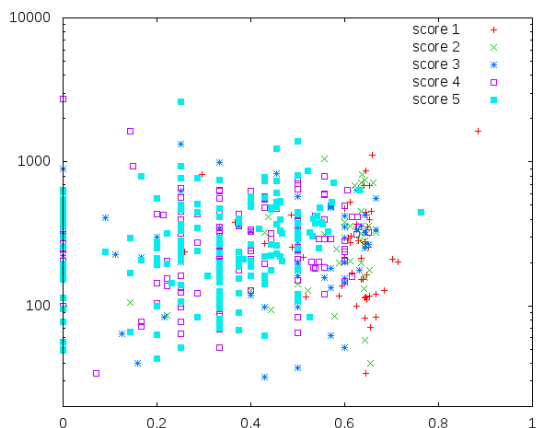


図2: 書評の分析 (ケース2: ベストセラー小説)

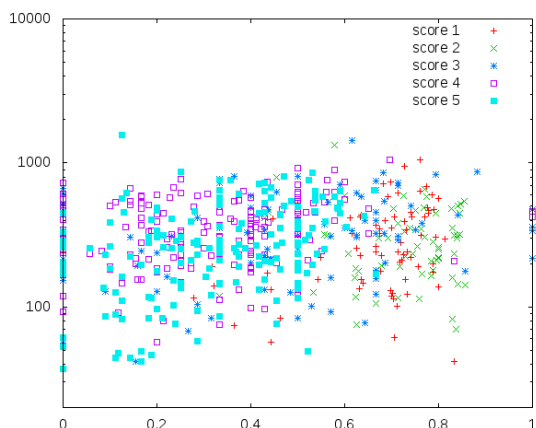


図3: 書評の分析 (ケース3: ミステリー小説)

全般的に有用度が下がっている。また星の数が低いコメントのほうが支持されており、星の数が高いも

のは共感を得にくいという傾向があるのではないかと、ということにも気付く。実際、これらのケースについて星5つのコメントと星1つのコメントに関する有用度の平均を検定すると、有意差があることが示される。この2作、いずれも内容とは違った文脈で話題が広がった作品であり、先の書評とはやや質が違っている。コアなファンによる書評以上にライトユーザによる書評が多く含まれると、全体としてはこのような傾向を見せるようである。

どうも集合痴となってしまったケース

ケース4: 次のグラフは、ある文学賞の大賞受賞者がイケメン俳優だったというレアなケース(図4)。完全に話題先行型で、この受賞作が出版されたときには、その内容に関してかなりの物議を醸した。こ

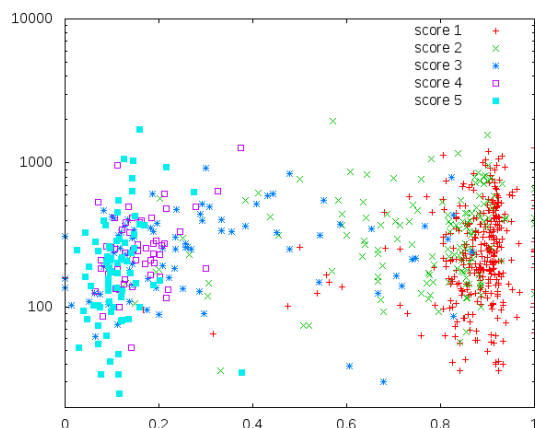


図4: 書評の分析 (ケース4: ある文学賞の大賞を受賞したイケメン俳優の作品)。集合痴となってしまったケース。

れまでに示したグラフと極端に異なることは一目瞭然である。星1つや星2つといった低評価の支持率が極めて高く、逆に星4つ、星5つといった高評価は軒並み反感を買っている。全体の評価結果に対する単純平均も低い。有用度で重みを付けて評価の平均をとると、悲惨な結果になってしまうだろう。

ケース5: 次は、不思議なベストセラーのケース(図5)。この作品も、先のグラフと比較的近いパターンとなっている。本作は、初版は自費出版として製作され、その文章も日本語としていろいろ問題があるという欠陥を抱えていたにも関わらず、ベストセラーとなり映画化までされたという作品である。グラフ全体の傾向はケース4(図4)と似ているが、

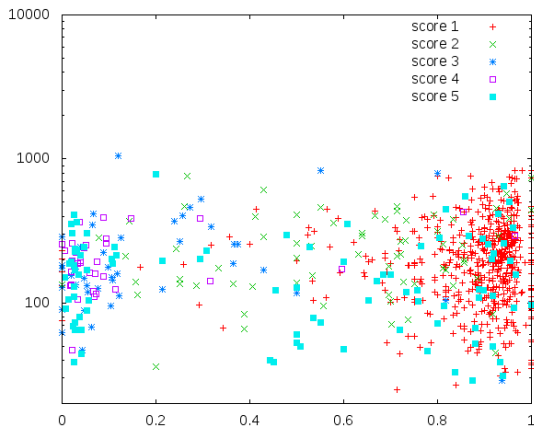


図 5: 書評の分析 (ケース 5: 不思議なベストセラー)

明らかに違う点は、星 5 つで有用度がわりと高いものが散見される点である。ところがこれらの星 5 つで有用度が比較的高いというレビューのほとんどが、書籍の内容に対するレビューではなく、レビューに関するメタレビュー、あるいは「本書はトイレトーパーとして最適でした」というようなウケを狙ったレビューである。その点には、十分に注意する必要があるだろう。次のケース 6 では、それがさらにエスカレートしている。

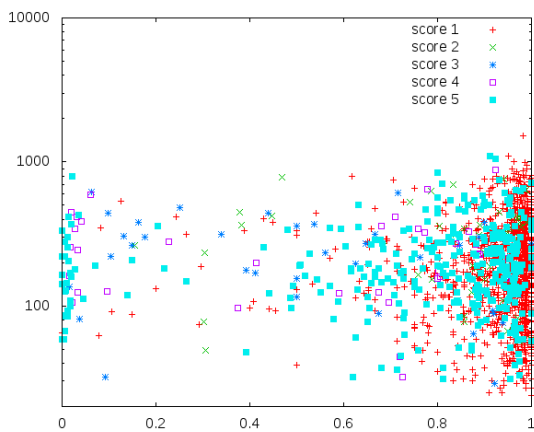


図 6: 書評の分析 (ケース 6: 極端な作品)

ケース 6: 本作は、1500 を越えるレビューがありながら、その評価は星 1 つまたは星 5 つに集中しているという極端な作品である (図 6)。1000 以上の評価が星 1 つとはいえ、ケース 4 (図 4)、ケース 5 (図 5) と異なり、星 5 つの評価も多くユーザによる支持を得ている特徴を持つ。しかしこの事実から、「内容がエキセントリックなので両極端に意見が割れているのかな」と安易に判断してはいけない。このグラフは、ケース 5 で指摘したようなメタレビューが

集中した結果である。このケースでは、完全にネットワークワーカの遊び場となってしまっており、いわゆる「縦読み⁴」という悪ふざけが蔓延した状態になっている。

集合痴には気を付けよう

本稿は、集合痴状態を否定するものではない。原因もなく掲示板が炎上したり、本稿でいう集合痴の状態になったりということはあり得ず、集合痴状態になるからには何がしかの要因が存在する。これまでの事例を振り返っても、炎上するケースというのは本人にも何らかの責任がある場合が多い。

本稿で指摘したいポイントは、CGM コンテンツを機械的に分析する際には気をつけましょうということである。集合痴状態になった書評を分析しても、その書籍が述べたい内容に関する有用な知見は得られないだろう (逆に、炎上した原因を含め、その書籍をとりまく文化的事例に対する分析はできるかもしれない)。CGM を頼る際には、集合痴か集合痴かを見極める能力を鍛え、落とし穴に落ちないように気を付けたい。

関連リンク：オンライン書評にまつわる様々な話題

1. 自作自演ばれる

http://tokyo.txt-nifty.com/fukublog/2004/02/post_129.html

2. 集合愚の世界

<http://t-c-management.seesaa.net/article/119176462.html>

3. 人はなぜつまらなかった本の書評を書くのか

<http://blogs.itmedia.co.jp/kyoko/2010/05/post-24eb.html>

4. 書評が広告になる時代

<http://blogs.itmedia.co.jp/mm21/2009/12/post-36ae.html>

5. 「ブログ炎上」の体験記

<http://www.nozomu.net/journal/000210.php>

6. 書評が消されてしまった話

<http://himadakadakara.blog80.fc2.com/blog-entry-61.html>

7. 書評の正しい見分け方

<http://sem-labo.net/blog/2009/10/04/0181/>

⁴<http://homepage2.nifty.com/Daiou2/Tateyomi.htm>
横書きの文章において、各行の頭文字をつなげるとメッセージが表れる文章を読ませる言葉遊び。既にこれはひとつの文化か。