

ユーザレビューの定量的比較の可視化 Visualization of Quantitative Comparison of User Reviews

小野 葵

Aoi Ono

法政大学情報科学部デジタルメディア学科

E-mail: aoi.ono.2t@stu.hosei.ac.jp

Abstract

Many user reviews of the evaluation of products are available from the Web, and are helpful to people in choosing which product to buy. However, it is difficult for people to read all such reviews because they are too many. This paper proposes a method for visualizing the user reviews of products by using a Treemap. It is possible for the method to visually present a large amount of information on reviews at one time. Processing user reviews to extract the words representing the evaluation of the products and then visualizing such words in a Treemap, the method makes it easy to compare the products. The paper provides the implementation of the method using morphological analysis to process user reviews of products obtained from the Web. Also, it presents the results of experiments on the visualization of the system using user reviews obtained from the Web. The experimental results visually show how each product is evaluated, what evaluation is common to both products, and what is specific to one of the products. Finally, the paper provides a discussion of the proposed method based on the experimental results.

1. はじめに

近年、Web 上に製品の評価をしたユーザレビューが数多く存在し、どの製品を購入すべきかを選択する際に参考になる。しかし、量が多いため、全てのレビューを読むことは難しい。例えば Web サイト、価格.com [1]の製品「ダイソン DC61 モーターヘッド」のページを閲覧すると、2015年12月22日時点で106件のレビューが投稿されている。この製品についての評価を知りたいときにこれらのレビューは有用であるが、全てを読んで参考にするには難しい。

本研究では製品を評価したユーザレビューを Treemap によって可視化する手法を提案する。この手法では取得した複数の製品のユーザレビューから評価を表す単語を抽出し、それを Treemap で可視化することで複数の製品の比較を容易にする。

この手法を応用し、Web 上にある製品を評価したユーザレビューを取得して、形態素解析によって単語を取得し可視化するシステムを作成した。また、Web 上のいくつかのユーザレビューを使用し、システムの可視化結果

に関する実験を行った。実験の結果、各製品がどのように評価されているか、どのような評価が両方の製品に共通するか、一方の製品に対してのみなされているのかを確認することが可能であった。本論文の最後に、実験結果に関する考察を述べる。

2. 関連研究

打田ら [2]はユーザレビュー上から評価を表すキーワードを抽出し時系列解析を行った。HKGraph で評価情報の視覚的把握支援を、多次元尺度法で月毎のレビューの評価項目によって定義されたレビュー間の類似度の可視化を行っている。HKGraph とは共起度情報に基づき、ユーザが入力した語句と関連性の高い語句をグラフ構造で可視化する手法である。

R. Sadana ら [3]は集合を1枚のレイヤに、集合内の各要素をレイヤ上の特定の位置にピクセルとして表示することで2値データの可視化を行った。更に複数枚のレイヤを重ね共通する部分のみを表示することで共通要素を見つけやすくなるとしている。

A. Lex ら [4]は集合の共通要素を可視化するための UpSet というシステムの研究、開発を行った。集合を列、共通要素を列内に表示する円のつながりで表している。共通要素の大きさや、それぞれの集合を共通要素毎に集合体として表示できることが特徴である。集合の属性値や要素の可視化も行うことができる。

M. Wang ら [5]は多重集合の共通部分を効率的に計算するための理論構築、手順設計を行った。更に多重集合の共通部分について可視化するためのスケーラブルな技術を開発した。この理論では集合や要素の数を任意に設定できるため幅広い用途に利用できると述べている。

3. Treemap

Treemap [6]とは、B. Shneiderman が1992年に提案した木構造の可視化手法である。一般的には木構造の各枝が長方形で表され、面積によってデータの大きさの認識、色によって項目の識別を行う。

図1は少人数グループの飲み物の嗜好についての Treemap である。大きな白枠が5種類の飲み物を表し、その中の色つきの長方形がグループ内の各個人を表している。この Treemap で面積は嗜好度の大きさ、色は個人を識別するために使われている。例えば飲み物の種類では「Coca Cola」の全体の面積が最も大きい。このためこのグループ内では「Coca Cola」が最も人気だと分かる。

また「Coca Cola」の中を見ると黄色で表された「Paul」が最も面積が大きい。これにより「Coca Cola」を好きなメンバーの中で「Paul」が最も好んでいることが分かる。更に同じ黄色の長方形は「Coca Cola light」と「Ice Tea」の2カ所にも確認できる。これから「Paul」は「Coca Cola」、「Coca Cola light」、「Ice Tea」の3種類の飲み物を好んでいることが分かる。



図1 Treemapの例

4. 提案手法

本論文ではユーザーレビューから取得した複数の製品の評価の定量的比較を可視化するために、Treemapを用いる手法を提案する。Treemapを使う理由は、数値的な大小関係が視覚的に分かりやすく、また階層構造を可視化できるためである。本研究では複数レビューの同時可視化を行うが、これに階層構造を使う事でレビューの評価の比較がしやすくなる。

本手法は以下の手順からなる。

1. まず複数の製品についてのユーザーレビューを取得し、評価を表している単語をそれぞれ抽出する。
2. 次に抽出した単語を、出現頻度を元にTreemapで可視化する。利用者は表示されたTreemapを見て製品がどのような評価をされているか比較することができる。

この手法では1つのTreemapに複数の製品の解析結果を表示する。ユーザーレビューから2つの製品AとBそれぞれのユーザーレビューを使用する場合、解析結果から取得できる単語は以下の3つに分類される。

- Aがより多く評価されているもの
- Bがより多く評価されているもの
- AとBで同数評価されているもの

解析結果から取得する単語の品詞は形容詞と形容動詞である。取得できる単語のうち、製品の評価を表す単語の品詞をその2つとし抽出する。

図2は1つのTreemapに複数の製品のレビューを可視化した結果の例である。Treemapは大きく3つの長方形で構成されており、製品Aがより多く評価されている単語、製品Bがより多く評価されている単語、製品AとBが同数評価されている単語に分かれて配置されている。このTreemapでは面積によって単語の出現頻度を表している。また色によって製品Aのレビューのみに現れる単語なのか、製品Bのみに現れる単語なのか、製品AとB両方に現れる単語なのかの識別を行う。青が製品Aに現れる単語、緑が製品Bに現れる単語、オレンジが製品A

とB両方に現れる単語を表す。青とオレンジ、緑とオレンジが重なっている単語は、製品A、B両方に現れる単語であるが、製品AまたはBのどちらか片方の出現頻度が高くなっている単語である。例えば図の右下に表示されている単語「大きい」は製品A、Bのレビュー両方に現れる単語であるが、製品Bのレビューでの出現頻度が高いため差分を製品Bの色である緑で表している。利用者はこの図を見て製品A、Bはそれぞれどのような評価を多くされているのか、2つの製品で同じ評価、異なる評価はどのようなものがあるのかを確認することができる。また、2色に分かれた単語を見たとき、色の割合を見てどちらにどの程度偏った評価であるのかが分かる。



図2 提案手法による可視化の例

本手法では各単語の出現頻度を用いて製品の評価を可視化している。これは各単語を元とし、出現頻度を重複度とする多重集合とみなすことができる。更に製品AとBのユーザーレビューから両方のレビューに現れる単語と一方のみに現れる単語を分けて可視化している。特に両方に現れる単語は、どちらの製品にどの程度偏っているかについても可視化している。これは多重集合の共通部分を可視化する一種のベン図とみなすことができる。

5. 実装

前節に述べた提案手法に基づき、製品のレビュー文を取得し形態素解析をしてその結果をTreemapで可視化するシステムを実装した。レビュー文を形態素解析し元となるデータを作成する処理にはPythonを、Treemapでの可視化にはJavaScriptを使った。また、Treemapの実装にはD3.jsというJavaScriptのライブラリを用いた。形態素解析にはYahoo! JAPANのWeb APIである日本語形態素解析[7]を利用している。このWeb APIは文章を渡すと形態素解析をした結果を返す。

まず、Pythonで作成したプログラムでユーザーレビューの解析を行う。製品A、Bの2つの製品のレビュー文を保存したテキストファイルを読み込み、Web APIに渡しそれぞれ形態素解析を行う。解析結果から形容詞、形容動詞の2つの品詞の単語を取得し、それらの語句をCSVファイルで保存する。CSVファイルには単語の出現頻度も同時に保存しておく。

次にJavaScriptで作成したプログラムでCSVファイルを読み込み、Treemapで可視化する。単語の出現頻度を比較し製品Aのほうが出現頻度の高い単語、製品Bのほうが出現頻度の高い単語、製品A、Bの出現頻度が同数の単語の3種類に分ける。分けた単語は木構造の3つの

枝に保存する。そして作成した木構造のデータを D3.js を用いて Treemap として可視化する。

6. 実験

実験に使うユーザレビューは価格.com に掲載されているものを使用した。まず価格.com からユーザレビュー文を抽出し、それを形態素解析する。そして作成したデータを Treemap で可視化する。製品レビューはレビュー数が 80 件以上のものを選んで使用している。

まず、同じカテゴリの 2 製品を使って Treemap を表示する実験を行った。図 3 は iRobot 製のロボット掃除機「ルンバ 780」と「ルンバ 770」のレビューを Treemap で可視化した結果である。この図ではオレンジで色付けされた部分が多く、2 つの製品レビューで同じ単語が多く現れている。その中で、「よい」「いい」「静か」「高い」などの評価は「ルンバ 780」のほうが多いことが分かる。

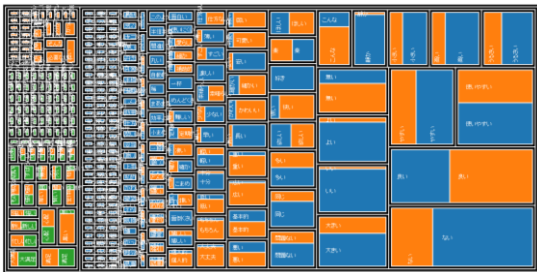


図 3 「ルンバ 780」(青)と「ルンバ 770」(緑)

図 4 は Dyson 製の掃除機である「DC61 モーターヘッド」と「Dyson Digital Slim DC45 モーターヘッド」の製品レビューについて可視化した結果である。「よい」「いい」「良い」の評価は「DC61 モーターヘッド」のほうが多い。また「使いやすい」の評価は「Dyson Digital Slim DC45 モーターヘッド」のほうが多い。

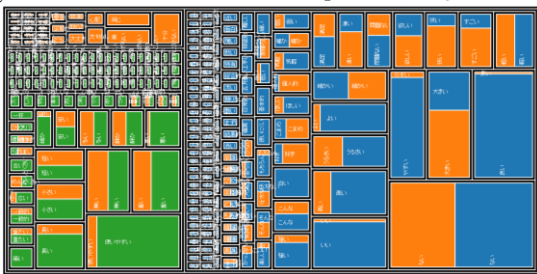


図 4 「DC61 モーターヘッド」(青)と「Dyson Digital Slim DC45 モーターヘッド」(緑)

図 5 は CANON 製のデジタルカメラ「PowerShot S120」と「PowerShot S110」のレビューを使用し可視化したものである。「PowerShot S110」は「良い」の評価が占める面積が大きく、評価が高いことが分かる。青の部分と緑の部分の配置が図 4 と異なるのは単語の量による。

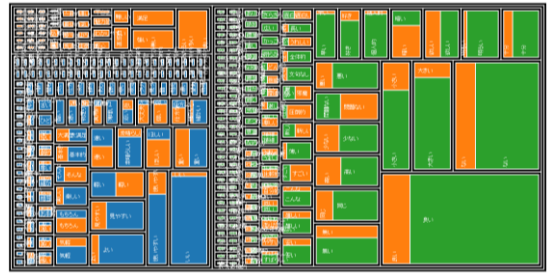


図 5 「PowerShot S120」(青)と「PowerShot S110」(緑)

以上の 3 つの例が示すように、この手法によって製品毎にどのような評価がされているか、また双方に含まれる評価と片方のみに含まれる評価はどのようなものがあるかが分かる。

次に別カテゴリの 2 製品を使って Treemap を表示する実験を行った。図 6 はカメラカテゴリの製品「PowerShot S120」と家電カテゴリの製品「ルンバ 770」を比較したものである。「良い」「見やすい」「やすい」など同じ単語が含まれているものがある。しかし「うるさい」「静か」などは「ルンバ 770」のエリアのみに含まれているのに対し「気軽」「明るい」「速い」などは「PowerShot S120」のエリアのみに含まれている単語であることが分かる。

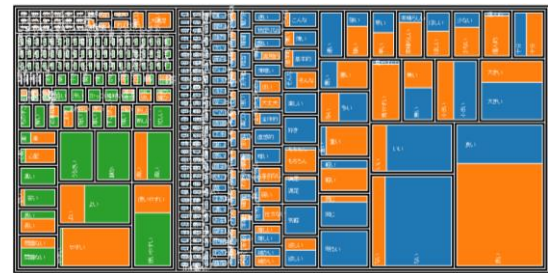


図 6 「PowerShot S120」(青)と「ルンバ 770」(緑)

次にカメラカテゴリの「PowerShot S120」と携帯電話カテゴリの「Xperia Z5 SO-01 H docomo」の 2 つの製品を使用して可視化する(図 7)。「問題ない」など両方に含まれる単語も多いが、「気軽」は「PowerShot S120」のエリアのみに、「熱い」は「Xperia Z5 SO-01 H docomo」のエリアのみに含まれていることが分かる。

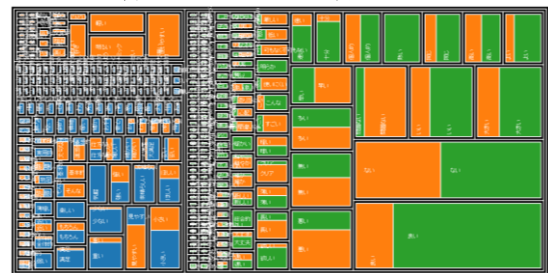


図 7 「PowerShot S120」(青)と「Xperia Z5 SO-01 H docomo」(緑)

最後に「ルンバ 780」と「Xperia Z5 SO-01 H docomo」の2つの製品レビューを使用して可視化する(図 8)。こちらも両方に含まれる単語は多いが、「静か」「こまめ」などは「ルンバ 780」のエリアのみに、「速い」「熱い」などは「Xperia Z5 SO-01 H docomo」のエリアのみに含まれることが読み取れる。

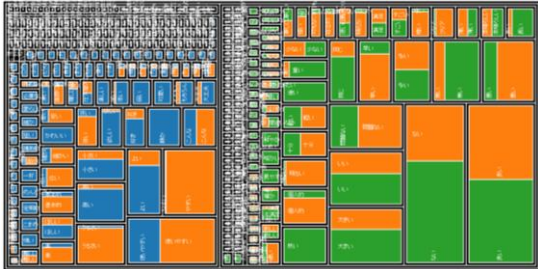


図 8 「ルンバ 780」(青)と
「Xperia Z5 SO-01 H docomo」(緑)

以上の 3 つの例が示すように、この手法によって製品に特徴的な評価を見つけることができる。

7. 議論

本手法で可視化することで製品毎にどのような評価がされているのかが 1 つの図を見ることで確認できるようになった。またそれぞれの製品毎の評価と双方に含まれる評価を、色分けをすることで、どのような評価がどの程度されているのかが分かるようになった。

問題点として、まず単語の選び方が適切かどうかという点が挙げられる。今回は形態素解析を行い形容詞、形容動詞のみを製品の評価を表す単語として抜き出し使用している。しかしそのため「ない」、「やすい」などの、その単語だけでは何を意味しているのか分からないものも含まれてしまっている。そしてそのような単語はレビュー中に多く現れるため Treemap で可視化したときに大きな面積を取ってしまう。これらの単語をストップワードとして事前に取り除くなどの工夫が必要だと考える。

また、Treemap に可視化したときにあまりにも面積が小さく読み取れない単語がある。これは出現頻度が少ないためである。レビュー数が多いと抽出される単語も多くなり 1 回しか現れない単語の数が多くなる。今回はレビュー数が 80 件以上の製品レビューを使用しているため端のほうの単語はほとんど読み取れなくなってしまっている。この問題点の解決法としてはまず図の表示を大きくする方法があるが、あまりにも大きい図は見にくいのではないかと考える。他の解決法としては一定数以下の単語を可視化時に切り捨てる方法もある。しかし、少数の意見が利用者にとって有益になることもあるので、必ずしも適切な方法とは言えない。本研究では導入しなかったが、長方形をクリックして拡大表示できる機能を追加する方法も考えられる。

8. おわりに

本論文ではユーザレビューから取得できる製品の評価情報を Treemap で可視化し比較する手法を提案した。提案手法の Treemap は面積で単語の出現頻度を表し、色で製品の評価に違いがあるのか、同じものが多いのかの識別を行った。また、そのためのシステムを Python と JavaScript を使用して作成した。システムはまず Python のプログラムでユーザレビューから製品の評価を行う単語を抽出し、JavaScript のプログラムで Treemap に可視化した。そしてそのシステムが有用であるか確認するため、Web 上のユーザレビューを使って可視化実験を行った。実験結果から、本手法は製品毎の評価が 1 つの図で確認でき、また色を分けて可視化することによってどの程度同じ評価がされているのか、異なった評価がされているのかが確認できることが分かった。また図の表示方法にはいくつか問題点があることが分かった。単語の選び方や出現頻度が低い単語の見やすさが挙げられる。単語の選び方についてはストップワードを設定するなどして改善する必要がある。図の見やすさについては出現頻度が低いものについては切り捨てたり、拡大表示をしたりするなどいくつか改善方法が考えられる。

文 献

- [1] "価格.com," Kakaku.com, Inc., [Online]. Available: <http://kakaku.com/>.
- [2] 打田裕樹, 吉川大弘, 古橋武, 平尾英司, 井口浩人, "Web ユーザレビューにおける評価情報の時系列変化の可視化," 日本知能情報ファジィ学会誌, vol. 22, no. 3, pp. 377-389, 2010.
- [3] R. Sadana, T. Major, A. Dove and J. Stasko, "OnSet: A Visualization Technique for Large-scale Binary Set Data," *IEEE Transactions on Visualization and Computer Graphics (IEEE InfoVis '14)*, vol. 20, no. 12, p. 1993-2002, 2014.
- [4] A. Lex, N. Gehlenborg, H. Strobel, R. Vuillemot and H. Pfister, "UpSet: Visualization of Intersecting Sets," *IEEE Transactions on Visualization and Computer Graphics (IEEE InfoVis '14)*, vol. 20, no. 12, pp. 1983-1992, 2014.
- [5] M. Wang, Y. Zhao and B. Zhang, "Efficient Test and Visualization of Multi-Set Intersections," *Scientific Reports*, vol. 5, pp. 16923:1-12, 2015.
- [6] B. Shneiderman and C. Plaisant, "Treemaps for space-constrained visualization of hierarchies," 1998. [Online]. Available: <http://www.cs.umd.edu/hcil/treemap-history/index.shtml>.
- [7] "テキスト解析:日本語形態素解析 - Yahoo!デベロッパーネットワーク," Yahoo Japan Corporation., [Online]. Available: <http://developer.yahoo.co.jp/webapi/jlp/ma/v1/parse.html>.