

科学研究費助成事業 研究成果報告書

令和元年6月20日現在

機関番号：32692

研究種目：若手研究(B)

研究期間：2016～2018

課題番号：16K16154

研究課題名(和文) 要点と評価を軸とした評価文書可視化に関する研究

研究課題名(英文) A Study on Visualization of Evaluation Documents Based on Gist and Evaluation

研究代表者

伏見 卓恭 (FUSHIMI, Takayasu)

東京工科大学・コンピュータサイエンス学部・助教

研究者番号：80755702

交付決定額(研究期間全体)：(直接経費) 2,700,000円

研究成果の概要(和文)：本研究では、ソーシャルメディアに投稿されたドキュメントを対象に、1)商品レビュー文の2部グラフ、2)ニュース記事などの時系列文書、3)学術論文の引用関係ネットワーク、4)商品に関する類似ツイートグラフ、5)トレンドワードバーストの5つの可視化手法を提案した。ソーシャルメディアに投稿される評価文書の多くは、ユーザの多様な意見だけでなく、商品に対するニーズが含まれている。評価観点やユーザニーズを軸として可視化することで、サービス提供側の計画策定時や消費者の商品購入時の意思決定に有用な情報を提示することができた。

研究成果の学術的意義や社会的意義

本研究の成果は、大量の評価文書を効果的かつ適切に可視化することでその全体像を把握し、さらに各評価文書へのアクセシビリティを向上させるための方法論である。ここで、効果的かつ適切な可視化とは、評価観点の類似性や評価極性、レビュアーの専門性などを考慮したスコアにより座標を決定することを意味する。この成果により、レビューサイトユーザの意思決定支援だけでなく、商品開発側の企業にも有用な情報発見が容易となることが期待される。さらに、これらの技術を一般化させ、インターネット上にある大量の文書データの整理、アクセシビリティ向上への貢献も十分期待できる。

研究成果の概要(英文)：In this research, for documents submitted to social media, we proposed the following visualization methods: 1) bipartite graphs of product review sentences, 2) time-series documents such as news articles, 3) citation relations network of academic papers, 4) similar tweets on products graph, 5) trend word burst. Many evaluation documents posted to social media include not only the diverse opinions of users but also the needs for products. By visualizing evaluation points and user needs as an axis, it is possible to present useful information for making decisions on the service provider side and for making decisions when consumers buy products.

研究分野：ウェブ情報学

キーワード：可視化 評価文書 時系列文書 極座標平面

様式 C-19、F-19-1、Z-19、CK-19（共通）

1. 研究開始当初の背景

本研究の主たる課題は、商品やサービスに対する評価文書（レビュー）を適切に可視化することで、対象商品の評価の全体像を直感的に把握でき、かつ所望の評価文書へのアクセシビリティを高めることにある。レビュー文を対象に評価表現・意見抽出に関する研究[1]や評価観点の抽出に関する研究[2]に端を発し、評価極性の定量化[3]、レビュー文の可視化[4]など多様な研究が展開されている。これらの研究の背景には、サイバー空間での購買行動においてユーザーの商品選択の手がかりが、既に商品を購入したユーザーによる評価文書であることが社会的な背景としてある。

しかしこれらの研究では、レビュアーの商品に対する知識、経験値などの専門性や、レビュアーごとの評点のばらつき傾向、偏り傾向が反映されていない。レビュー文には、ユーザーの異なる背景や文脈が反映されており、それらを加味することは他のユーザーの商品選択に際し有用である。商品に対する専門性の類似の概念として、あるブランドの商品を繰り返し購入することを指すブランドロイヤリティ[5]がある。ブランドロイヤリティが高いユーザーのレビューは、同一ブランド他商品との比較などが盛り込まれていたり、長年の知識・経験が反映されていることが期待できる。また、レビュアーごとの評点のばらつき、偏りに関する研究も広く展開されている[6]。本研究では、レビュアーの専門性や評点の偏りを定量化するこれらの技術を積極的に用いて、可視化結果に反映させることを目指す。レビューの可視化に関する既存研究では、離散的に評価観点を抽出し、タグクラウドや棒グラフ、レーダーチャート形式でプロットする手法が多く見られる[4,7]。抽出した評価観点の数には恣意性があったり、その数が多いほど解釈困難な結果となる。また、隣接する軸の関連性については言及されていない。よって、ユーザーの商品選択において有用な可視化結果を出力するには限界があると言える。

〔参考文献〕

- [1] Turney, P.D.: “Thumbs Up or Thumbs Down? Semantic Orientation Applied to Unsupervised Classification of Reviews,” ACL, 2002.
- [2] Hu, M. and Liu, B.: “Modeling Online Reviews with Multi-grain Topic Models,” KDD04, 2004.
- [3] Scaffidi, C. et al.: “Red Opal: Product-Feature Scoring from Reviews,” EC, 2007.
- [4] Oelke, D. et al.: “Visual Opinion Analysis of Customer Feedback Data,” VAST, 2009.
- [5] 守口 剛: “潜在クラス・ロジット・モデルを利用したロイヤリティ・セグメンテーション,” オペレーションズ・リサーチ, 2003.
- [6] Salganik, M. J. et al.: “Experimental study of inequality and unpredictability in an artificial cultural market”, SCIENCE, 2006.
- [7] Wu, Y. et al.: “OpinionSeer: Interactive Visualization of Hotel Customer Feedback,” IEEE Transactions on Visualization and Computer Graphics, 2010.

2. 研究の目的

本研究の基本課題は、レビューサイトや EC サイトなどに投稿される大量の評価文書（レビュー）を、要点と評価により可視化する手法の確立である。評価はレビュアーごとの評点のばらつき度、偏り度、商品に対する専門性を反映させたスコアを算出し利用する。サイバースペースでの購買行動においてユーザの商品選択の手がかりとして、既に商品を購入したユーザによるレビューがメインとなる。口コミなどの評価文書は有用な情報の宝庫であり、集合知として整理、活用することで、商品選択時の重要な参考情報になると期待できる。ゆえに、本研究の目的は、各商品に対する評価の全体構造把握や、ユーザ所望の評価文書発見支援を見据えた可視化手法を確立することである。

3. 研究の方法

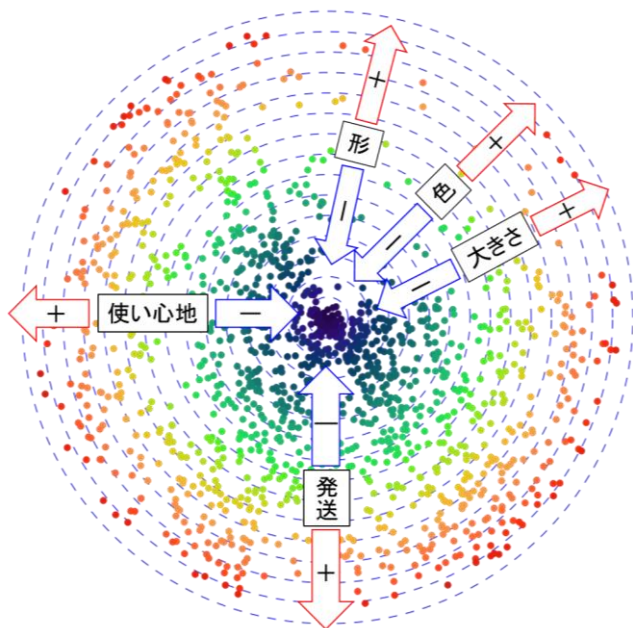
上述した本研究の目的を達成するために、以下の4つの課題を遂行する。

- (1) 要点と評価を考慮した評価文書の可視化手法の確立
- (2) レビューアーの専門性や評点の偏り傾向の定量化、および可視化手法への統合方法確立
- (3) 可視化結果の評価法確立と可視化結果から得られる傾向分析
- (4) 可視化結果へアノテーション（注釈）を付与する手法の確立、および有効性の評価

(1) 要点と評価を考慮した評価文書の可視化手法の確立

まず最初に、ある商品に関する全レビューを対象とした可視化法を提案する。レビューなどの評価文書集合を $D = \{i, \dots\}$ 、単語集合を $W = \{j, \dots\}$ 、とし、各レビューを Bag of Words の単語頻度ベクトル a_i で表現する。 a_{ij} は、 a_i の j 番目の要素であり、レビュー i に単語 j が出現した回数である。文書と単語それぞれをノード、その出現関係を重み付きリンクとした二部グラフとして扱う。文書ノード、単語ノードはそれぞれ半径 r_1 、 r_2 の円上に配置し、同心円状に可視化する。具体的には、重み行列 $A = [a_{ij}]$ に対して中心化を施し、中心化重み行列 $B = [b_{ij}]$ を得る。目的関数 $J(X, Y) = \sum_{i=1}^{|D|} \sum_{j=1}^{|W|} b_{ij} \frac{x_i^T}{r_1} \cdot \frac{y_j}{r_2}$ を最大化するようにノード座標行列 $X = (x_1, \dots, x_{|D|})^T$ および $Y = (y_1, \dots, y_{|W|})^T$ を反復的に求める。但し、 x_i, y_j は、文書ノード i 、単語ノード j の布置座標ベクトルを表し、 $\|x_i\| = r_1$ 、 $\|y_j\| = r_2$ である。この目的

関数において 2 つのベクトルの内積は $\frac{x_i^T}{r_1} \cdot \frac{y_j}{r_2} = \cos(\theta_{ij})$ であり、隣接するノード（文書 i に出現する単語 j ）どうしがその重みに応じて原点から見て同じ方向に配置されることによって、 $J(X, Y)$ は最大化される。すなわち、同じ単語が出現する文書どうし、および、同じ文書に出現する単語どうしを近くに配置する。また、ベクトル群 Y を固定すれば、ベクトル群 X の最適配置が求まり、 X を固定すれば Y の最適配置が求まるため、反復解法により効率的に解を求める。（左図参照）



(2) レビュー어의専門性や評点の偏り傾向の定量化, および可視化手法への統合方法確立

各レビュー어의評点ばらつき, 偏り度や専門性などを反映させたスコアを算出し, スコアに基づきレビュー文をプロットする. レビュー어의専門性算出に関しては, 商品カテゴリを対象に, レビュー어의カテゴリ選択モデルを提案する. 提案モデルにレビュー어ごとの実レビュー群を適用した際の尤度により, レビュー어의カテゴリ専門性を算出する. 本研究では, このモデルに対しカテゴリ情報以外に, 自然言語処理技術を用いて抽出する商品特徴量も用いる. さらに, テキスト部分における専門用語などを考慮することで拡張し, レビュー어의専門性を表す指標を提案する. レビュー어의評点の偏り傾向算出に関しては, ジニ係数を算出し, ユーザのある商品への評点とそのユーザにとってどの程度ポジティブな点数かネガティブな点数であるかを算出する. これは, ユーザによって異なる評点の意味を正規化する役割を果たす. これらの値を評点に加味することで算出したスコアにより, 可視化における半径を定める.

(3) 可視化結果の評価法確立と可視化結果から得られる傾向分析

提案可視化手法は, 動径と偏角により定まる極座標平面上に評価文書をプロットする. 偏角は評価文書の要点や評価観点を, 動径(半径)はレビュー어의専門性や評点の偏りを考慮したスコアをそれぞれ表す. 偏角の評価として, トピックモデルを用いて類似トピックの評価文書が同一方向にあるかを定量的に評価することを検討している. 動径の評価として, その構成要素である評点, レビュー어의評点偏り具合, レビュー어의専門性それぞれに関して, 評価極性との相関などにより定量的に評価する.

(4) 可視化結果へアノテーション(注釈)を付与する手法の確立, および有効性の評価

上述の(1)~(3)で確立した可視化技術により, 評価文書群をその要約とスコアによりプロットした. この技術を使用するには, 可視化結果のどのあたりに(原点から見てどの方向に)どのようなレビューが集まっているかを明示的に示さなければならない. 本研究では, 円周上にプロットされた評価文書群に対し, vonMises 分布の意味で密集した部分文書集合を抽出する. そして, 各部分文書集合に統計的有意に出現する特徴量(単語)により, アノテーションを付与する.

4. 研究成果

ソーシャルメディアに投稿されたレビュー文を対象に, スコアとコメント文から抽出する評価観点を軸とし, 極座標平面上にレビューをプロットする手法を確立した. コメント文と文中に現れる単語からなる二部グラフを構築し, 提案した目的関数を最大にするように各ノード(コメント文と単語)の最適配置を決定する. これにより, 類似の単語から構成される, すなわち, 類似の観点で評価されたコメント文は, 原点から見て同一方向にプロットされるようになる. さらに, スコアにより原点からの距離(半径)を決定する. 加えて, 近傍に位置するノードの共通特徴を用いて可視化結果へアノテーションを付与する手法を提案した[雑誌論文(3)]. この手法を時系列文書[雑誌論文(1)]や学術文献の引用ネットワーク[学会発表(13)], 商品に関する類似ツイートグラフ[学会発表(3), (9)], トレンドワードバースト[学会発表(6)]の可視化にも応用した. 本研究の成果は, 複数の雑誌論文や国際会議に採録されているため, 国内外におけるある一定のインパクトがあると言える. 今後の展望として, 各文書の特徴をより顕著に表現するためのベクトル化技術と, 時々刻々と投稿される時系列文書をオンラインで可視化する技術を導入することで, より効果的・効率的な可視化手法を実現することが望まれる.

5. 主な発表論文等

[雑誌論文] (計 3 件)

- (1) 伏見 卓恭, 佐藤 哲司, ``時系列文書に対するトピックフォレストの構築と構造解析'', 日本データベース学会和文論文誌, Vol.16-J, No.1, pp.1--8, 2018.3. 【査読あり】
- (2) 伏見 卓恭, 斉藤 和巳, 風間 一洋, ``メトリック空間における複数カテゴリに属するハイブリッドオブジェクト抽出法の提案'', 情報処理学会論文誌 ネットワーク科学特集号, Vol.58, No.6, pp.1258--1267, 2017.6. 【査読あり】
- (3) 伏見 卓恭, 佐藤 哲司, 斉藤 和巳, 風間 一洋, ``ネットワーク上での特徴量分布を考慮したアノテーション付与法'', 情報処理学会論文誌 ネットワーク科学特集号, Vol.58, No.6, pp.1246--1257, 2017.6. 【査読あり】

[学会発表] (計 19 件)

- (1) Takayasu Fushimi, and Ryota Mori, ``High-Speed Clustering of Regional Photos Using Representative Photos of Different Regions,'', Proc of the 2018 IEEE/WIC/ACM Int'l Conf. on Web Intelligence (WI2018), pp.520--527, 2018/12/04. 【査読あり】
- (2) Takayasu Fushimi, and Mio Ono, ``Modeling Opinion Formation by Incorporating Users Expertise and Friendshipness,'', Proceedings of the 20th International Conference on Information Integration and Web-based Applications & Services (iiWAS2018), pp.103--110, 2018/11/20. 【査読あり】
- (3) 菅野 健一, 伏見 卓恭, ``類似ツイートグラフ構築のための類似度閾値決定法'', 第120回数理モデル化と問題解決研究発表会 (MPS120), 2018/09/26.
- (4) 伏見 卓恭, 斉藤 和巳, 池田 哲夫, 風間 一洋, ``グラフ連結性に基づく多様体上での頑健なクラスタリング手法'', 第120回数理モデル化と問題解決研究発表会 (MPS120), 2018/09/26.
- (5) 毛利 亮太, 伏見 卓恭, ``他地域の代表写真を利用した高速クラスタリング手法の改善'', 第120回数理モデル化と問題解決研究発表会 (MPS120), 2018/09/26.
- (6) 松井 直大, 伏見 卓恭, ``極座標可視化手法を用いたトレンドワードのバースト検出'', 第120回数理モデル化と問題解決研究発表会 (MPS120), 2018/09/26.
- (7) 小野 水緒, 伏見 卓恭, ``ユーザ特性を考慮した Voter Model における高影響度ノードの分析'', 第120回数理モデル化と問題解決研究発表会 (MPS120), 2018/09/25.
- (8) 毛利 亮太, 伏見 卓恭, ``クラスタ代表写真に基づく地域間の共通特徴の抽出'', 第17回情報科学技術フォーラム (FIT2018), 2018/09/21.
- (9) 菅野 健一, 伏見 卓恭, ``類似ツイートグラフに基づくユーザニーズの可視化手法'', 第17回情報科学技術フォーラム (FIT2018), 2018/09/21.
- (10) 毛利 亮太, 伏見 卓恭, ``他地域の代表写真を利用した地域別写真の高速クラスタリング手法の提案'', マルチメディア, 分散, 協調とモバイルシンポジウム 2018 (DICOMO2018), 2018/07/06.
- (11) 瀧上 翔生, 伏見 卓恭, ``多様なレシピ群を形成する便利食材群の推薦'', マルチメディア, 分散, 協調とモバイルシンポジウム 2018 (DICOMO2018), 2018/07/06.
- (12) 小野 水緒, 伏見 卓恭, ``ユーザの専門度を考慮した意見形成モデルの提案'', マルチメディア, 分散, 協調とモバイルシンポジウム 2018 (DICOMO2018), 2018/07/06.
- (13) Takayasu Fushimi, Tetsuji Satoh, and Noriko Kando, ``Dynamic Visualization of Citation Networks and Detection of Influential Node Addition,'', Proceedings of the 9th International Conference on Complex Networks (CompleNet2018), pp.291--302, 2018/03/07. 【査読あり】
- (14) 小林 滉河, 野崎 祐里, 伏見 卓恭, 佐藤 哲司, ``消費者の検索行動に基づいた EC サイトの動的なカテゴリ生成手法の提案'', 第10回 Web とデータベースに関するフォーラム (WebDBf2017), 2017/09/18.
- (15) 伏見 卓恭, 斉藤 和巳, 池田 哲夫, 風間 一洋, ``機能クラスタ抽出高速化の一検討'', 第14回ネットワーク生態学シンポジウム (NetEco14), 2017/08/21.
- (16) Takayasu Fushimi, and Tetsuji Satoh, ``Constructing and Visualizing Topic Forest for Text Stream,'', Proceedings of the 2017 IEEE/WIC/ACM International Conference on Web Intelligence (WI2017), pp.10--17, 2017/08/23. 【査読あり】
- (17) 稲福 和史, 伏見 卓恭, 佐藤 哲司, ``EC サイトにおける購買行動のモデル化と成長分析'', マルチメディア, 分散, 協調とモバイルシンポジウム 2016 (DICOMO2016), 2017/06/29.
- (18) Takayasu Fushimi, Kazumi Saito, Tetsuo Ikeda, and Kazuhiro Kazama, ``Accelerating Greedy K-medoids Clustering Algorithm with L1 Distance by Pivot Generation,'', Proceedings of the 23rd International Symposium on Methodologies for Intelligent Systems (ISMIS2017), pp.87--96, 2017/06/27. 【査読あり】
- (19) 伏見 卓恭, 斉藤 和巳, 風間 一洋, ``一般化ピボットによる L1 距離に基づく K-medoids クラスタリングの高速化'', 2017 年度人工知能学会全国大会 (第31回) (JSAI2017), 2017/05/25.

6. 研究組織

※科研費による研究は、研究者の自覚と責任において実施するものです。そのため、研究の実施や研究成果の公表等については、国の要請等に基づくものではなく、その研究成果に関する見解や責任は、研究者個人に帰属されます。