

話題に対する感情俯瞰グラフ

上岡由征^{†1} 河合由起子^{†1} 若宮翔子^{†1} 熊本忠彦^{†2} 張建偉^{†3} 白石優旗^{†3}
^{†1} 京都産業大学 ^{†2} 千葉工業大学 ^{†3} 筑波技術大学

1 はじめに

Web 検索結果にはニュースサイトや企業サイト等の公共性の高いニュース記事だけでなく、ブログやマイクログログといった個人の嗜好性の高い記事が混在する。本研究では、特定の話題に関するニュース記事とツイートの書き手の感情の相違を俯瞰可能な可視化システムを構築する。我々はこれまで、感情辞書 [1] を用いて Web 検索結果のタイトルとスニペットの感情値を算出し、感情値に基づく検索システムを提案、実装してきた [2]。このシステムは、ユーザが検索キーワードを入力すると、検索 API を用いて Web ページのタイトルとスニペットを取得して感情値を算出し、感情分布を提示する。これにより、検索キーワードに対する感情分布を把握することができる。一方、特定のニュースサイトのニュース記事に対するコメントや Twitter 上に投稿されたニュースに関するコメントを対象とした感情分析に関する研究 [3, 4] も行われている。しかしながら、これらの既存研究では同一話題に関するニュース記事とツイートにおける感情の相違は考慮していない。

本論文では、検索キーワードに関するニュース記事とツイートの感情値を感情辞書を用いて算出し、感情俯瞰グラフにプロットすることで、ニュース記事とツイートの感情分布を比較可能なシステムを提案する。また、2014 年に流行した話題を検索キーワードとした評価実験を行い、マスメディアと大衆の話題に対する感情分布の推移に関する知見を考察する。

2 システム概要

図 1 に提案システムの流れを示す。まず、ユーザが入力した検索キーワード (例えば、「衆議院選挙」など) を用いて、ニュース記事とツイートをそれぞれ取得する。ニュース記事は、タイトルとスニペットを取得する。ツイートに関しては、ユーザの意見が述べられている有効なツイートを抽出するために、リツイートのみでユーザの意見が述べられてないツイートを除く。次

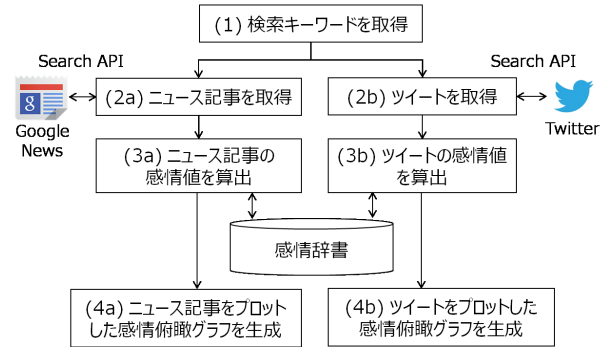


図 1: システムの流れ

に、既存の感情辞書を用いて、各テキストに含まれる語の感情値を抽出し、ニュース記事とツイートの感情値をそれぞれ算出する。感情値は、3つの感情軸 (軸 1:「楽しい ⇔ 悲しい」、軸 2:「うれしい ⇔ 怒り」、軸 3:「のどか ⇔ 緊迫」) の値の組として抽出される。最後に、ニュース記事とツイートをそれぞれの感情値に基づき 3次元の感情俯瞰グラフにプロットする。

3 実験

図 2 と図 3 に検索キーワード「衆議院選挙」で感情俯瞰グラフを用いた感情分布の可視化例を示す。図 2(a) のプロットはニュース記事、(b) のプロットはツイートである。各グラフは 2本の感情軸 (x軸, y軸) からなっており、プロットの座標で 2軸の感情値を、プロットの色で 3軸目の感情値を表している。感情俯瞰グラフを用いることにより、図 3 のように検索キーワードに関するニュース記事とツイートの感情分布を時間経過とともに提示することができる。

次に、話題に対するニュース記事とツイートの感情が各軸のどのあたりに分布しているか、またその広がりを確認した。表 1 に実験に用いた 10個の検索キーワードに対するニュース記事とツイートの各軸の感情値の平均と標準偏差の結果を示す。表 1 より、全ての検索キーワードの全ての軸において、ツイートの感情値の標準偏差よりもニュース記事の感情値の標準偏差の方が大きかった。また、平均を比較したところ、ニュース記事の感情値の平均とツイートの感情値の平均との差が大きく (標準偏差 9.0 以上)、反する感情が抽出されたものがあつた (表中の太字)。例えば、「羽生結弦」という検索キーワードの場合、ニュース記事は

A Proposal of Comprehensive Sentiment Graphs for a Topic Using Tweets and News

^{†1} Yoshiyuki UEOKA ^{†1} Yukiko KAWAI ^{†1} Shoko WAKAMIYA
^{†2} Tadahiko KUMAMOTO ^{†3} Jianwei ZHANG ^{†3} Yuhki SHIRAISHI

^{†1} Kyoto Sangyo University
^{†2} Chiba Institute of Technology
^{†3} Tsukuba University of Technology

表 1: 話題に対するニュース記事とツイートの感情分布の比較

検索 キーワード	ニュース記事						ツイート					
	平均			標準偏差			平均			標準偏差		
	軸 1	軸 2	軸 3	軸 1	軸 2	軸 3	軸 1	軸 2	軸 3	軸 1	軸 2	軸 3
アベノミクス	49.76	52.41	51.23	20.08	22.25	22.91	55.90	59.30	45.45	9.06	11.39	11.42
エボラ	52.00	53.80	45.13	20.75	23.74	24.05	54.79	58.07	47.29	13.11	13.96	14.50
増税	55.49	41.30	43.96	19.51	26.82	24.07	59.76	64.94	54.13	10.26	17.14	14.07
高倉健	53.61	65.10	59.90	16.97	14.55	21.06	56.49	60.52	47.83	8.32	9.63	9.92
羽生結弦	58.95	72.30	56.99	17.39	17.79	19.51	58.46	61.75	45.41	7.72	8.59	10.80
本田圭佑	58.73	68.80	53.22	17.32	18.88	25.24	59.79	59.62	47.04	11.41	13.27	12.36
田中将大	56.79	72.64	52.55	19.26	14.85	22.70	56.83	58.95	45.80	6.58	8.38	7.29
錦織圭	48.88	67.14	60.69	23.81	24.99	24.90	58.76	62.35	48.56	9.30	11.63	9.90
妖怪ウォッチ	53.08	63.51	65.67	19.80	17.11	21.14	56.83	59.22	47.50	12.17	14.47	15.65
アナと雪の女王	41.29	58.28	62.64	24.04	26.32	20.23	58.60	60.55	49.94	15.87	15.16	14.50

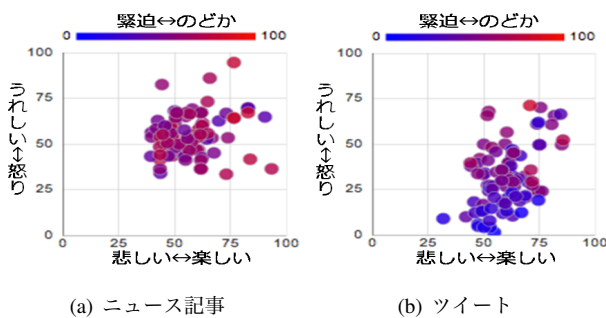


図 2: 感情俯瞰グラフの例

「楽しい」、「うれしい」、「のどか」な感情を中心に分布していたが、ツイートは「楽しい」、「うれしい」、「緊迫」した感情を中心に分布していた。また、「増税」の場合、ニュース記事は「楽しい」、「怒り」、「緊迫」した感情を中心に分布していたが、ツイートは「楽しい」、「うれしい」、「のどか」な感情を中心に分布していた。

4 おわりに

本研究では、特定の話題に関するニュース記事とツイートの書き手の感情分布の相違を俯瞰可能な可視化システムを構築した。そのために、感情辞書を用いてニュース記事とツイートの感情値をそれぞれ算出し、感情俯瞰グラフを生成する手法を提案した。実験では、流行話題に関するニュース記事とツイートを取得し、それぞれの感情分布の差異について考察した。

今後の課題として、政治やスポーツ等のジャンルにおける話題に対するニュース記事とツイートの感情の差異パターンを検出し、提案システムの有効性検証を行う予定である。また、短文のテキストであるツイートの感情を適切に抽出することは難しい。複数のツイートに分けて同一のトピックについての意見を発信しているユーザも多いため、文脈を考慮したツイート取得手法についても検討する予定である。

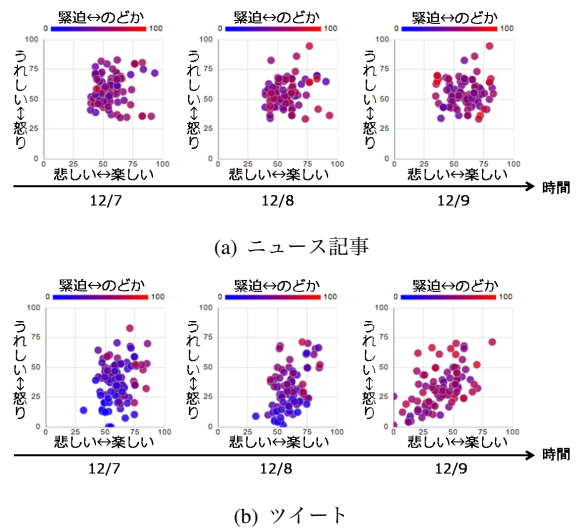


図 3: 感情分布の時間変化の可視化

謝辞

本研究の一部は、総務省戦略的情報通信研究開発制度 (SCOPE) および JSPS 科研費 24780248, 26330347, 26870090 の助成を受けたものである。ここに記して謝意を表す。

参考文献

- [1] 熊本忠彦, 河合由起子, 田中克己, “新聞記事を対象とするテキスト印象マイニング手法の設計と評価”, 電子情報通信学会論文誌, J94-D(3), pp. 540-548, 2011.
- [2] Jianwei Zhang, Yukiko Kawai, and Tadahiko Kumamoto, “Extracting Similar and Opposite News Websites Based on Sentiment Analysis,” In Proc. of 2012 International Conference on Industrial and Intelligent Information (ICIII 2012), pp. 24-29, 2012.
- [3] 但馬康宏, “コメントとしてつぶやかれた短文の感情推定”, 電子情報通信学会技術研究報告. NLC, 言語理解とコミュニケーション 112(196), pp. 37-40, 2012.
- [4] 内藤和宏, 榎堀優, 梶田将司, 間瀬健二, “Twitter コメントに含まれる感情語がイベント印象に与える影響の評価”, 情報処理学会インタラクティブ, pp. 871-876, 2012.