

# ニュースサイトにおける話題およびセンチメント差異 提示システムの検討

松本 好史<sup>†</sup> 張 建偉<sup>‡</sup> 河合 由起子<sup>‡</sup> 中島 伸介<sup>‡</sup>

京都産業大学<sup>‡</sup>

## 1. はじめに

近年, Web の普及により多くのニュースサイトが様々なサービスを提供するようになった. 我々はこれまでニュース記事の発信者 (ニュースサイト, 地方, 国など) のセンチメント差異を抽出し, 地図上に表示するセンチメントマップシステム [1, 2] を開発してきた. センチメントマップシステムでは, 話題に関する発信者の全体的なセンチメント傾向をニュースサイトレベル, 地方レベルあるいは国レベルで把握できる. これにより, 立場の違いによる多様な情報発信が明らかとなった. しかしながら, ある話題に対して, 一つのニュースサイト内に複数の関連トピックの記事が書かれていることは一般的である. たとえば, 「鳩山首相」に対して, 「子ども手当」に関連する記事は「承認」のセンチメントを持つ一方で, 「献金」に関連する記事は「拒否」のセンチメントを持つ.

そこで, 本研究では, 検索キーワードとその関連語, さらにそれらを含む記事のセンチメントの 3 つの観点に基づき, 各ニュースサイトの差異を提示するシステムを提案する. 具体的には, まず, 検索キーワードを含む記事から関連語を抽出する. 次にユーザは任意の関連語かニュースサイトまたはその両方を指定する. 最後に, ユーザの指定に応じたセンチメント値を, 感情軸ごとにグラフとして表示する. 提案システムを使用することにより, ユーザはニュースサイト間, 関連語間, およびニュース記事間のセンチメント差異を把握できる.

## 2. システムの概要

### 2.1 関連語の抽出方法

- 検索キーワードを含む一定期間内の一定数の記事を取得する.
- A) で取得した記事において形態素解析を行い, 「名詞-非自立-一般」と「名詞-接尾-一般名詞-接尾-人名」を除く名詞の単語を抽出する.
- B) で抽出した単語の  $tf \cdot idf$  値を式(1)より求め, 値が高い 10 単語を関連語とする.

A System for Presenting the Distinction of Topics and Sentiments in News Websites

Yoshifumi Matsumoto, Jianwei Zhang, Yukiko Kawai, Shinsuke Nakajima,  
Kyoto Sangyo University

$$tf \cdot idf(w, Pi) = \frac{C(w, Pi)}{N(Pi)} \times \log \frac{N}{N(w)} \quad (1)$$

$C(w, Pi)$ : ニュース記事  $Pi$  で抽出された単語  $w$  の出現回数

$N(Pi)$ : ニュース記事  $Pi$  で抽出された全単語の出現回数

$N$ : 収集した記事数

$N(w)$ : 単語  $w$  を含む記事数

### 2.2 センチメント値の算出方法

各単語のセンチメント値は, 先行研究 [1, 2] で開発したセンチメント辞書から求める. センチメント辞書は, 各単語がもつセンチメント値を示す辞書である (表 1). 本研究の特徴の一つとして, ポジティブとネガティブなセンチメントだけではなく, より人間の感情に近いとされる 4 次元のセンチメントを分析する. 単語のセンチメント値は 0~1 の値をとる. たとえば, 単語「蘇生」の感情軸 1 「明るい⇔暗い」におけるセンチメント値は 0.910 であり, 単語「蘇生」が「明るい」センチメントに偏ることを示す. 記事のセンチメント値は, 記事内の単語のセンチメント値の平均である.

表 1 センチメント辞書の一例

単語	感情軸 1 明るい⇔暗い	感情軸 2 承認⇔拒否	感情軸 3 緩和⇔緊張	感情軸 4 怒り⇔恐れ
挑戦する	0.618	0.687	0.752	0.500
衝突する	0.344	0.353	0.315	0.529
死亡	0.280	0.358	0.260	0.364
脱線	0.310	0.546	0.403	0.291
蘇生	0.910	0.521	0.429	0.000
豊富だ	0.597	0.676	0.761	0.466

### 2.3 関連語およびニュースサイトごとのセンチメント値の提示

ユーザが関連語を指定した場合は, 指定した関連語のセンチメント値をニュースサイトごとに表示し, その関連語に対するニュースサイト間のセンチメントの差異を比較することができる (図 1). ユーザがニュースサイトを指定した場合は, 指定したニュースサイト内の各関連語に対するセンチメント値を関連語ごとに表示し, そのニュースサイト内の各関連語に関するセンチメントの差異を比較することができる (図 2). また, 関連語とニュースサイトを両方指定した場合は, 指定したニュースサイト内の指定した

関連語に関する記事のセンチメント値を記事ごとに表示し、各記事のセンチメントの差異を比較することができる(図 3)。なお、感情軸 1 は「明るい⇔暗い」、感情軸 2 は「承認⇔拒否」、感情軸 3 は「緩和⇔緊張」、感情軸 4 は「恐れ⇔怒り」をそれぞれ表している。

### 3. プロトタイプシステムによる実行結果

今回は以下の 4 つのパターンで実験を行った。

- 1) 検索語「巨人」関連語「シリーズ」
- 2) 検索語「雪」関連語「冬」
- 3) 検索語「景気」関連語「経済」
- 4) 検索語「オリンピック」関連語「招致」

1, 2 はセンチメントの差異が大きいと予想され、3, 4 はセンチメントの差異が小さいと予想されるパターンである。

実験結果は予想通りパターン 1, 2 ではサイト間のセンチメント値の差異が大きく、パターン 3, 4 ではサイト間のセンチメント値の差異はあまり見られなかった。パターン 1 で北海道新聞の感情軸 3 が他のサイトと比べて小さいのは、地元チームが日本シリーズで巨人に敗れたのが原因であると考えられ、パターン 2 で沖縄タイムスと他のサイトとのセンチメント値の差異が大きいのは、気候が大きく異なることが原因と考えられる。また、感情軸 2, 4 はパターンに関わらず、サイト間のセンチメントの差異はあまり見られなかった。

### 4. まとめ

本論文では、「検索キーワード」、「関連語」、「発信者」の 3 つの観点からセンチメントの相違を提示するシステムを提案した。実験では、3 つの観点による相違をグラフで確認することができた。今後、各記事や発信者のセンチメントの相違に基づいた信頼性支援を行う予定である。また、今回の実験ではあまり機能していない感情軸があったので、話題に応じた感情軸の変更も検討したい。

### 謝辞

本研究の一部は、独立行政法人情報通信研究機構による委託研究「電気通信サービスにおける情報信憑性検証技術に関する研究開発」の成果であり、ここに記して謝意を表す。

### 参考文献

- [1] 藤田裕介, 河合由紀子, 熊本忠彦, 張建偉, 田中克己: 「地域性に基づく発信者の観点可視化システムの提案」, Web とデータベースに関するフォーラム (WebDB Forum2008), 2008 年 12 月。
- [2] Jianwei Zhang, Yukiko Kawai, Tadahiko Kumamoto, and Katsumi Tanaka. "A Novel Visualization Method for Distinction of Web News Sentiment". In Proc. Tenth International Conference on Web Information Systems Engineering (WISE 2009), LNCS 5802, pp. 181-194, October 2009.

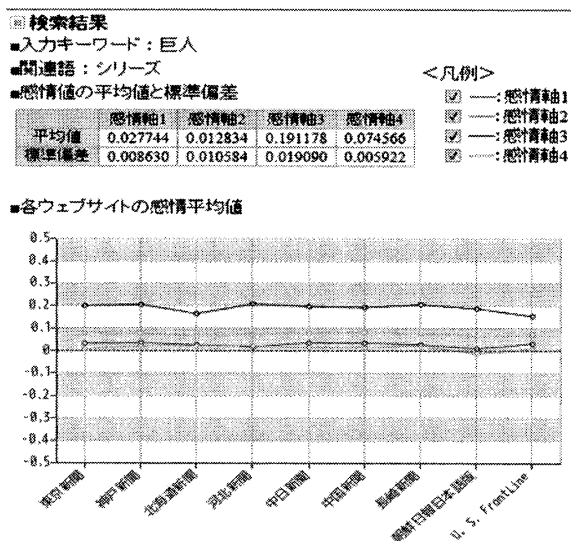


図 1 ニュースサイト間のセンチメントの差異

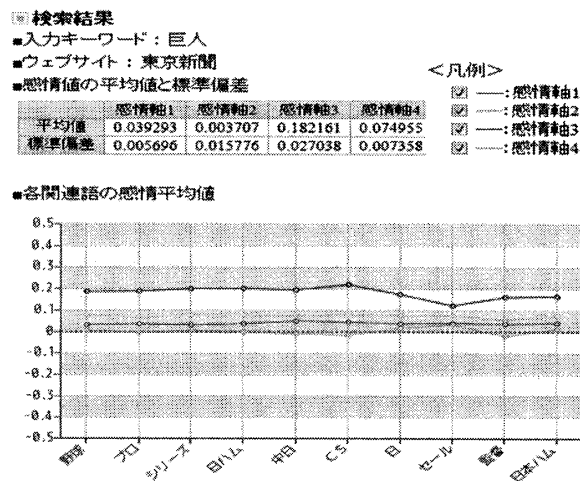


図 2 関連語間のセンチメントの差異

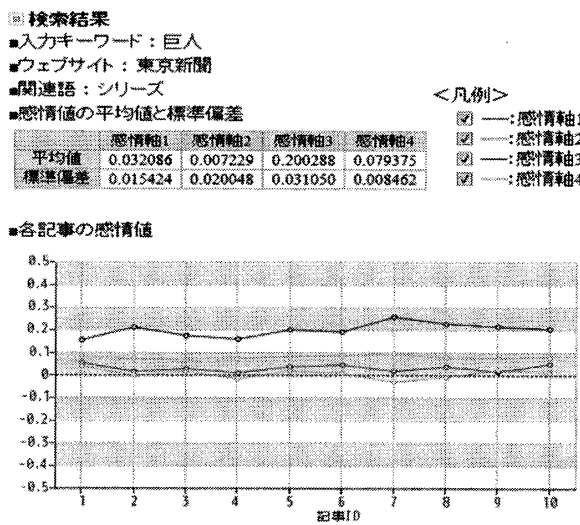


図 3 ニュース記事間のセンチメントの差異