

感情に基づく対極性感情検索手法の提案

Impression-based Search Technique for Obtaining Opposite Web Pages

南 勝利[†] 河合 由起子[†] 張 建偉[‡] 白石 優旗[†] 熊本 忠彦[§]

Katsutoshi Minami Yukiko Kawai Jianwei Zhang Yuhki Shiraishi Tadahiko Kumamoto

1. はじめに

近年, Web の膨大な情報の中から効率良くまた効果的に情報を発見するための技術として, 汎用の検索エンジンにより得られた検索結果を分類し提示する研究や, 再検索の手間を軽減するインタフェースに関する研究が活発に行われている [1][2]. 基本的な検索結果の分類手法では, 単語の出現頻度や単語の共起関係から特徴的な単語を抽出し, それら抽出した単語の検索結果を分類を用いて行う研究がある [3][4]. 再検索の手間を軽減するインタフェースでは, 追加の検索キーワードをユーザが入力する必要がなく, 例えば検索結果で提示されるタイトルとスニペット (概要) に表示されている単語をマウスで選択することで, 「強調 (AND 検索)」「削除 (NOT 検索)」を容易にする研究がある [5][6].

しかしながら, これらの手法は, 出現する単語のメタ情報 (ユーザに与える印象や情景, 感情など) を十分に考慮しているとは言えない. 一方で, 文書中の反対語に基づき文脈に依存した評価表現の感情極性を抽出する研究など, 文章や単語に対する「感情」というメタ情報を抽出する研究も活発に行われている [7]~[9].

我々はこれまで検索結果から, 我々が開発した感情辞書 [10] を用いて検索キーワードより複数の感情値という新たな特徴量を複数個抽出し, それら感情値の組み合わせに基づき, リランキング可能なシステムを提案し, 実装評価してきた [11]. 提案システムでは, まず, 各検索結果に含まれるタイトルとスニペットからなるテキストデータからその感情値を抽出する. 感情値は, 楽しい⇔悲しい, 嬉しい⇔怒り, のどか⇔緊迫という対極性を持つもので, これらを 3 軸の感情軸として定義する. 次に, 各感情軸ごとに全検索結果から平均値を算出し, 検索キーワードに対する感情値としてユーザに提示する (図 1). また, 検索結果のリランキングは提示された各感情軸の感情値を変更することで, 変更された感情値に基づいて, 検索結果はリランキングされる.

この感情軸により, ユーザは検索キーワードに対する

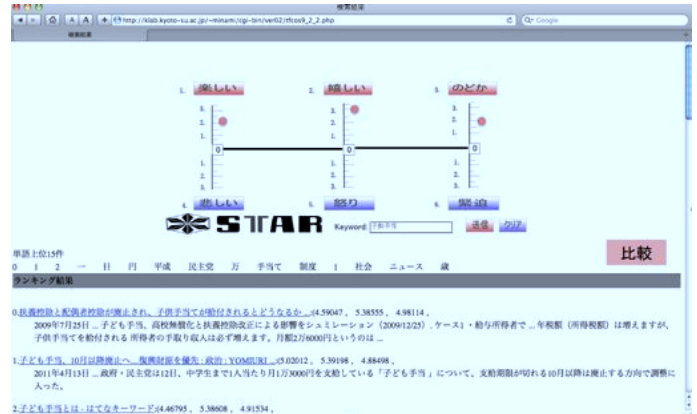


図 1: 検索語と対極性の高い検索結果の提示例

メジャーな感情 (例えば嬉しいといった感情値が高い) を把握することができる. しかしながら, 感情軸の値を対極側に変更することで反対の感情を持つ意見 (ページ) も発見できるが, そもそも対象としている検索結果上位の多くには反対の感情を含む検索結果が少ないため, 対極意見となるマイナーな感情を含むページをほとんど発見できない. また, マイナーなページとして検索結果の下位を対象とする方法では, 検索キーワードとの関連性が低くなってしまう.

本研究では, 我々が開発したページ閲覧者が抱く感情の期待値である「感情値」を用い, 先行研究として行ってきた検索キーワードに対して世の中が抱くメジャーな感情を抱かせるページの発見 [11] のみならず, 対極となるマイナーな感情を抱かせるページの発見を可能にする「対極性感情検索手法」を新たに提案する. 具体的には, まず検索結果から得られる検索キーワードの感情値から距離最大のページを選出し, 次にその中から重要度の高い関連語を抽出し, 距離最大の単語を対極性関連語として抽出する. 最後に, 対極性関連語を補間語とし再検索し, 対極性感情検索結果として提示する.

本論文の構成は, まず第 2 章でシステムの概要を述べ, 第 3 章で感情値算出手法および対極性関連語抽出手法について述べる. 次に, 第 4 章で評価実験では, 感情値検

[†]京都産業大学 Kyoto Sangyo University

[‡]筑波技術大学 Tsukuba University of Technology

[§]千葉工業大学 Chiba Institute of Technology

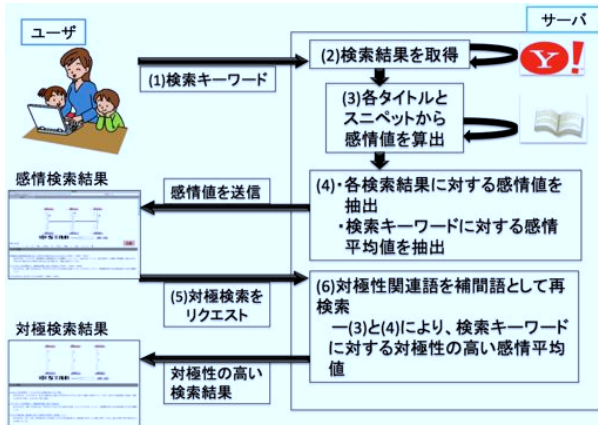


図 2: 全体の流れの図

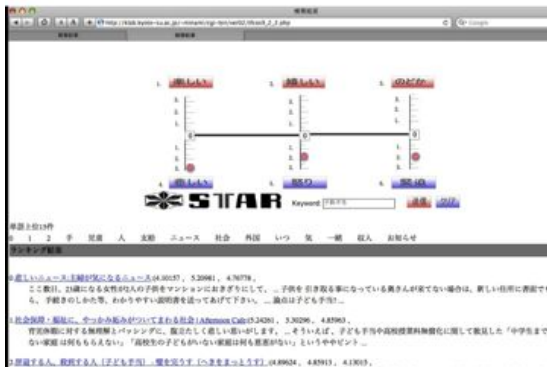


図 3: 検索語と対極性の高い検索結果の提示例

索および対極検索を考察し、今後の課題を述べ、最後に 5 章でまとめとする。

2. システムの概要

本システムの処理の流れを図 2 に示す。

1. ユーザから検索キーワード (例えば、子供手当) が与えられる。
2. サーバ側では、検索キーワードを受信すると、Yahoo! の検索 API を用いて検索結果を取得する。検索結果ページの取得方法は、処理コストと精度のトレードオフにより検索結果 M 件の中から N 件をサンプリングする。その後、取得した検索結果の各ページのタイトルと概要であるスニペットを取得する。
3. 検索結果のタイトルとスニペットに対して、感情辞書を用いて感情値を算出する。検索キーワードに対する感情値は、各検索結果の感情値の平均値とする。
4. 3 軸の感情平均値を検索キーワードに対する感情値としてグラフを検索結果とともに提示する (図 1)。なお、検索結果より抽出した関連語も提示する。
5. ユーザが比較ボタンをクリックすると、検索キーワードと感情平均値がサーバに送られ、対極感情検索をリクエストする。なお、グラフの感情軸の感情値を調整・設定したりランキングも可能である。
6. サーバはユーザから検索キーワードと 3 軸の感情平均値を受信すると、感情平均値を要素とするベクトルと各検索結果の感情値を要素とするベクトルとのユークリッド距離を算出する。次に、このユークリッド距離を基に検索結果を選別し、選別した検索結果から対極性関連語を抽出する。この対極性関連語を補間語として検索キーワードを用いて手順 3 と 4 により再検索し、対極性感情検索結果をユーザへ提示する (図 3)。

3. 感情値抽出と対極性関連語抽出

各検索結果ページに対する感情値の算出法と、対極性の高い検索結果を取得するために用いる補間語となる対極性関連語の抽出法を説明する。

3.1 感情値の算出

感情値の算出には感情辞書を使用する [10]。感情辞書の作成方法を以下に述べる。まず、感情軸を構成する感情語群を設定する (表 1)。また、 IW_L と IW_R とは感情語群を表し IW_L は表 1 の「楽しい、楽しむ、楽しみだ、楽しげだ」、「嬉しい、喜ばしい、喜ぶ」、「のどかだ、和やかだ、素朴だ、安心だ」とし、 IW_R は「悲しい、悲しむ、悲しみだ、悲しげだ」、「怒る、憤る、激怒する」、「緊迫する、不気味だ、不安だ、恐れる」とする。次に、6 年分の朝日新聞、読売新聞の記事データ (2002~2007) から感情語を含む記事を抽出し、記事に含まれる感情語群 IW_L に属する感情語と感情語群 IW_R に属する感情語の数を比較し、 IW_L の数が多い記事の集合を S_L (記事数を N_L)、 IW_R の数が多い記事の集合を S_R (記事数を N_R) とする。このとき、ある単語 w の記事集合 S_L における出現記事数を $N_L(w)$ 、記事集合 S_R における出現記事数を $N_R(w)$ とすると、それぞれの補正済み条件付確率は、

表 1: 感情語群

感情軸	感情語
楽しい ⇔悲しい	楽しい, 楽しむ, 楽しみだ, 楽しげだ 悲しい, 悲しむ, 悲しみだ, 悲しげだ
嬉しい ⇔怒り	嬉しい, 喜ばしい, 喜ぶ 怒る, 憤る, 激怒する
のどか ⇔緊迫	のどかだ, 和やかだ, 素朴だ, 安心だ 緊迫する, 不気味だ, 不安だ, 恐れる

表 2: 感情辞書一例

単語 1 ⇔ 0	感情軸 1 楽しい⇔悲しい	感情軸 2 嬉しい⇔怒り	感情軸 3 のどか⇔緊迫
初受賞	0.862	1.000	0.808
クッキング	1.000	0.653	0.881
ひなまつり	0.847	1.000	0.977
偽装	0.245	0.075	0.297
死刑だ	0.013	0.028	0.000
拘束する	0.059	0.103	0.000

$$P_L(w) = \frac{N_L(w)}{N_L} \quad (1)$$

$$P_R(w) = \frac{N_R(w)}{N_R} \quad (2)$$

と表される。この $P_L(w)$ と $P_R(w)$ を用いて、単語 w の感情値 $s(w)$ を次のような式で表す。

$$s(w) = \frac{P_L(w) * weight_L}{P_L(w) * weight_L + P_R(w) * weight_R} \quad (3)$$

$$weight_L = \log_{10} N_L \quad (4)$$

$$weight_R = \log_{10} N_R \quad (5)$$

また、これら動作は 1 軸のものなので各軸分 3 回繰り返す。さらに、我々は、感情辞書を用いて算出されるテキストの感情値を説明変数、そのテキストに対し被験者 100 人が与えた 7 段階評価値の平均値を目的変数とする回帰分析を感情軸ごとに行い、その結果得られた回帰式 (3 次関数あるいは 5 次関数) を用いて、感情辞書により算出されたテキストの感情値を補正するという手法を提

案している。なお、感情値 $s(w)$ は 0~1 の値をとる。ここで、 $s(w)$ が 1 に近い際は「楽しい, 嬉しい, のどか」という感情を表し、0 に近い際は「悲しい, 怒り, 緊迫」という感情を表す。感情辞書の一例を表 2 に示す。この例では、「初受賞」という単語の「楽しい⇔悲しい」という感情軸の感情は 0.862 であり、「楽しい」感情を表す。「偽装」という単語の「嬉しい⇔怒り」という感情軸の感情は 0.075 であり、「怒り」感情を表す。

ページの感情値は、そのタイトルとスニペットに出現した各単語の感情の平均で算出する。ここで、ページの感情値は 0~1 の値をとる様に設定されているが、ユーザの利便性と対称性を考え、「楽しい」「嬉しい」「のどか」が強いつきは 3、「悲しい」「怒り」「緊迫」が強いつきは -3 という設計にした。そのため、換算値 = 6 * 算出値 - 3 によりスケールリングを行った。また、図 2 に示す通り、チャート表示の際は、各感情軸の値は 0~3 の絶対値表記とした。

3.2 対極性関連語抽出

前節より、ページに対する 3 つの感情値を抽出した。これらの検索結果の感情値から各軸ごとに平均値を算出し、検索キーワードに対する感情平均値とする。

対極性関連語の抽出は、下記の手順とする。

1. 検索キーワードに対する感情平均値ベクトル $V_q=(v_{q1}, v_{q2}, v_{q3})$ とし、検索結果における各ページの感情値ベクトル $V_p=(v_{p1}, v_{p2}, v_{p3})$ に対して、対極性を判定する。判定は、各軸に対して行い、符号が逆転している場合に対極性ありとする。例えば、 v_{qi} が正 (負) の時 v_{pi} が負 (正) であれば対極性有とする。
2. 対極性が有ると判定された場合、その軸のユークリッド距離である $\sqrt{(v_{qi} - v_{pi})^2}$ を算出する。対極性が無い軸は 0 とする。
3. ユークリッド距離が最大となるページを対象とし、Yahoo!API の特徴語抽出により関連語を抽出する [12]、なお、tf-idf による特徴語抽出も行ったが Yahoo!API の関連語とほぼ同程度の抽出となったため、処理コストを考慮し、本稿では Yahoo!API の SCORE 値が 40 以上の関連語を用いる。
4. 抽出した関連語から重複していないもので、感情値が感情平均値とのユークリッド距離が最大のものを対極性関連語とする。

表 3: 検索キーワードと対極性関連語

ア) Positive Word	(1)	(2)	(3)	(4)
検索キーワード	人種のるつぼ	東京大学	ネットカフェ	エイプリルフール
対極性関連語	混血	消費	派遣	自粛
イ) Negative Word	(5)	(6)	(7)	(8)
検索キーワード	記者会見	橋下徹	裁判員裁判	東京電力
対極性関連語	キャンペーン	スケジュール	手順	エネルギー
ウ) Positive Negative 混合	(9)	(10)	(11)	(12)
検索キーワード	老後	就職活動	バブル経済	防犯カメラ
対極性関連語	チカラ	メディア	働	こだわり

抽出した対極性関連語を補間語(サブキーワード)として検索キーワードと組み合わせて再検索する。再検索結果を取得し、タイトルとスニペット、感情辞書を用いて各検索結果の感情値および感情平均値を算出し、ユーザへ対極性感情検索結果として提示する。

4. 評価実験

実装した提案システムについての評価実験を以下の通り行う。提案システムは、PHP Version 5.2.6, FLASH ver 8.0 で開発し、検索エンジンには、Yahoo!検索 Web API を用いた。対極性関連語の抽出に使用する検索結果ページ数は、先行研究 [13] より、上位約 2000 件中からサンプリングした 86 件が網羅性が高く、かつ関連性が高いことを確認しており、本実験でもこれを採用した。

対極性感情検索の評価は、検索キーワードに対する感情平均値と対極性感情検索結果から算出された感情平均値を比較することにより行う。実験は、2011 年 12 月上旬に実施し、12 個の検索キーワードを用いた(表 3)。このうち時事キーワードとして「橋下徹」、「裁判員裁判」、「東京電力」、定性的なキーワードとして、「人種のるつぼ」、「防犯カメラ」、「エイプリルフール」とした。また、最初の感情値検索において 3 軸の感情平均値に基づき次の 3 つに分類された。

- ア) 感情平均値が 3 軸全て正 (Positive)
(楽しい, 嬉しい, のどか)
- イ) 感情平均値が 3 軸全て負 (Negative)
(悲しい, 怒り, 緊迫)
- ウ) 正と負の両方が含まれるもの
(楽しい, 怒り, 緊迫), (悲しい, 怒り, のどか) 等

なお、表 3 には、各検索キーワードに対して提案システムが抽出した対極性関連語も示す。対極性感情検索では、この対極性関連語が補間語として用いられた。

4.1 3つの感情軸ごとの対極性感情検索結果

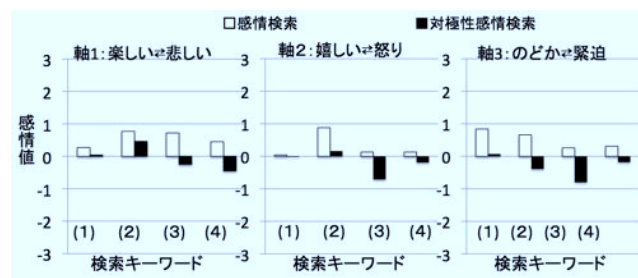


図 4: 感情検索がすべて Positive な場合の比較

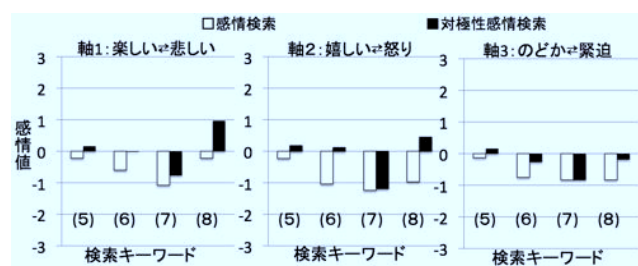


図 5: 感情検索がすべて Negative な場合の比較

図 4 から図 6 に、検索キーワードにおける感情検索結果と対極性感情検索結果による感情平均値を、各軸ごとに詳細比較した結果を示す。縦軸は感情平均値を示しており、換算値の 3 から -3 の値をとり、正が各々「楽しい」、「嬉しい」、「のどか」を表し、負が「悲しい」、「怒り」

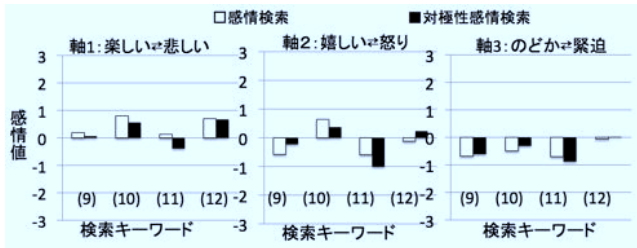


図 6: 感情検索が Positive, Negative 両方を含む場合の比較

り、「緊迫」を表わす。横軸は表 3 の検索キーワードとなる。ここで、グラフの左棒（白）が感情検索結果の感情平均値を表し、右棒（黒）が対極性感情検索の感情平均値を表す。

図 4 より、すべての感情平均値が Positive な場合では、すべてのキーワードの全軸において対極性感情検索の結果の感情値が小さくなり、Negative 側に傾いたことが示されている。図 5 の、感情検索結果が Negative なキーワードの場合では、(7) 裁判員裁判の軸 3 のみで同値となった以外は全て逆側に傾いたことが示されている。図 6 の Positive と Negative 混合型のキーワードの場合では、(11) バブル経済の軸 2 と軸 3 以外で全て逆側に傾いたことが示されている。また、検索されたページの内容を比較すると、(4) 「エイプリルフール」を例とすると、感情検索結果の上位は、「エイプリルフール」ネタサイト 7 選 - ロケットニュース 24」といった「楽しい」感情の内容の記事であった。一方、対極性感情検索結果の上位には「サイトでのエイプリルフール「ネタ」今年も自粛の企業も：J-CAST ニュース」といった「悲しい」内容の記事となった。

4.2 全体の対極性感情検索結果

ア) Positive のみ、イ) Negative のみ、ウ) Positive と Negative の混合の 3 つの分類において、感情検索結果と対極性感情検索結果による感情平均値を比較した。その結果、感情検索がア) Positive に分類された検索キーワードでは、全体の平均で Negative の対極側に 0.64 傾いた。イ) Negative に分類された検索キーワードでは、全体の平均で対極側の Positive に 0.54 傾いた。ウ) Positive と Negative 混合の場合は、各軸ごとの差を算出し、各単語ごとに平均を算出して、全体の平均値を比較した。その結果、各単語ごとの平均が Positive だったものが Negative 側に、また Negative だったものが Positive 側の対極側に傾いた平均値は 0.14 であった。また、検索

されたページの内容を比較すると (8) 「東京電力」の感情検索結果は「カスタマーセンター検索 | 東京電力」となりある程度 Negative な感情を抱かせる内容の記事が上位であったが、対極性感情検索では、「東京電力 呷取りサービス編 明日をひらくエネルギー - YouTube」といった Positive な感情の内容が上位を占めた。

4.3 考察と今後の課題

図 4 から図 6 の結果より、いくつかの軸で感情値が対極側に傾かないものもあったが、概ね対極側に傾いた。また、3 つに分類した検索キーワード群に対する全体平均は、若干ではあるが 3 つともすべて対極側に傾いた。以上のことから、感情検索結果に対して対極性のある感情を含むページを多く検索できたことが示された。

今後は、再検索結果より得られた検索結果を用いてより対極性の高い関連語を抽出し、対極性を高める手法を検討する。また、時間的な変化も考慮した提示手法についても検討する。

5. まとめ

本研究では、検索キーワードに対するメジャーな感情値を提示すると同時に、それらの感情と対極性の高い所謂マイナーな検索を可能にする対極性感情検索システムを提案した。評価実験の結果より、検索キーワードに対する感情平均値に対して対極性をもつ検索結果の取得が確認できた。今後は再検索を繰り返すことで、より対極性の高いページの発見手法を検討する。

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) の一環として実装されたものである。ここに記して謝意を表す。

参考文献

- [1] H. Zeng, Q. He, Z. Chen, W. Ma and J. Ma: "Learning to cluster web search results", Proc of SIGIR2007. pp. 210-217 (2004)
- [2] 関 隆宏, 和多 太樹, 山田 泰寛, 廣川 佐千男, 検索支援と分析のための多面的検索システム, 電子情報通信学会 第 19 回データ工学ワークショップ (DEWS2007), E1-2, 2007.
- [3] 野田 武史, 大島 裕明, 手塚 太郎, 小山 聡, 田中 克己, Web 検索結果のクラスタリングに用いる話題語の質問キーワードからの自動抽出, 電子情報通信学

会 第 18 回データ工学ワークショップ (DEWS2006),
2C-i8, 2006.

- [4] Clusty the Clustering Engine
<http://clusty.jp/>
- [5] Rerank. jp <http://rerank.jp/>
- [6] 吉田 大我, 小山 聡, 中村 聡史, 田中 克己, Web
検索結果におけるキーワード出現相関の可視化と対
話的な質問変換, 第 18 回データ工学ワークショ
ップ (DEWS2007), C7-2, (2007)
- [7] 酒井 義和, 荒木 健治, 反対語を利用した文脈依存
評価表現の感情極性判定, 電子情報通信学会論文誌,
Vol.J93-D, No.9, pp.1778-1789, 電子情報通信学会
2010
- [8] 阿部健一, 藤本悠, 大原剛三, ユーザーが受信メー
ルから抱く感情に合わせた音楽推薦システム, 第 4
回データ工学と情報マネジメントに関するフォーラ
ム (DEIM2011), No. A9-1, 2012.
- [9] K. H. Lin, C. Yang, H. Chen, Emotion Classifi-
cation of Online News Articles from the Reader's
Perspective, Proc. of the 2008 IEEE/WIC/ACM
International Conference on Web Intelligence and
Intelligent Agent Technology, Vol. 1, pp. 220-226,
2008.
- [10] 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,
論文誌 (D), Vol.J94-D, No.3, pp. 540-548, 2011
年 3 月.
- [11] 早崎麻耶, 前川大地, 張建偉, 河合由起子, 熊本忠
彦, 感情値を用いた検索結果のリランキング手法の
提案, 電子情報通信学会 第 3 回データ工学と情報
マネジメントに関するフォーラム (DEIM) 2011.
- [12] Yahoo!キーフレーズ抽出 API
[http://developer.yahoo.co.jp/webapi/jlp/keyphra
se/v1/extract.html](http://developer.yahoo.co.jp/webapi/jlp/keyphrase/v1/extract.html)
- [13] 荒深康夫, 河合由起子, 張建偉, 熊本忠彦, レーダー
チャートを用いた柔軟なリランキング手法の実装,
情報処理学会インタラクシオン 2010