

# 感情値を用いた検索結果のリランキング手法の提案

早崎 麻耶<sup>†</sup> 前川 大地<sup>†</sup> 張 建偉<sup>†</sup> 河合由起子<sup>†</sup> 熊本 忠彦<sup>††</sup>

<sup>†</sup> 京都産業大学理学部コンピュータ科学科 〒603-8555 京都市北区上賀茂本山

<sup>††</sup> 千葉工業大学情報科学科情報ネットワーク学科 〒275-0016 千葉県習志野市津田沼 2-17-1

E-mail: <sup>†</sup>{g738174,g738363,zjw,kawai}@cc.kyoto-su.ac.jp, <sup>††</sup>kumamoto@net.it-chiba.ac.jp

あらまし 近年、検索結果を分類提示する研究や、再検索の手間を軽減をするリランキングに関する研究開発が活発に行われている。本研究では、検索キーワードに対する感情を値に変換したもの(センチメント値)を提示し、それらの各値をインタラクティブに修正することでセンチメントに基づく、リランキングを実現することを目的とする。具体的には、検索キーワードに対するセンチメント値は、検索結果  $N$  件に含まれるタイトルとスニペットから感情辞書を用いて、3軸の各々が対極性を持つ各センチメント値を抽出し、それらの平均値とする。次に、ユーザが修正した3軸のセンチメント値と、各検索結果のセンチメント値に対して類似性判定を行い、類似性の高い順にリランキングする。本手法により、検索結果の表層のテキストデータには含まれない深層部分のセンチメント値の算出や対極性を利用した網羅性の高いリランキングが可能となる。本稿では、センチメントに基づくリランキング手法について提案する。

キーワード センチメント, リランキング

## 1. はじめに

近年、検索エンジンの結果を利用し、検索結果を分類する研究や、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多く行われている [1] [2]。基本的な検索結果の分類手法は、検索キーワードを補完する単語の出現頻度や共起関係を抽出し、それらの単語ごとに検索結果を分類する手法が一般的である [3] [4]。再検索の手間を軽減する手法では、検索キーワードを入力ボックスにユーザが記入する必要がなく、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してやると、「強調(AND検索)」「削除(NOT検索)」が容易にできる [5]。これらの研究開発により、検索キーワードを考慮あるいは入力する手間が軽減できる。しかしながら、これらの再検索手法は、出現する単語のメタ情報間の関連性を考慮しているとは言えない。

本研究では検索結果のテキストデータから、センチメントをメタ情報として抽出し、それらのセンチメントに基づいたリランキングシステムを提案する。提案システムでは、まず、検索結果から検索キーワードに対するセンチメントをユーザに提示する。ユーザに提示されるセンチメントは、楽しい 悲しい, 嬉しい 怒り, のどか 緊迫という対極性をもつ3軸となる。ユーザは、提示された複数のセンチメントとその値を変更することができる。変更されたセンチメント値に基づき、検索結果はリランキングされる。これにより、複数のセンチメントを組み合わせた検索ができる。また、対極性をもつため、提示された検索結果のセンチメントと反対のセンチメントを用いたリランキングも可能である。

以下、本論文では、2章でシステムの概要を述べ、3章で実装による検討を行い、4章でまとめとする。

## 2. システム設計

### 2.1 システムの概要

本システムの処理流れを図1に示す。

- (1) ユーザから検索キーワードが与えられる。
- (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、検索結果  $M$  件の中から一定の間隔で126件をサンプリングをする。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、我々がこれまでに開発した感情辞書を用いてセンチメント値を算出する。検索キーワードに対するセンチメント値は、各検索結果のセンチメント値の平均で算出される。
- (4) 3軸のセンチメントと検索キーワードに対するセンチメント値を、グラフの項目と各項目の初期値として、検索結果とともにユーザに提示する(図2)
- (5) ユーザは、取得した検索結果に対して、関心を持つセンチメントを調整・設定できる。例えば、検索キーワードに対する楽しいセンチメントのページに興味を持つ場合は、ユーザは「楽しい」軸の値を高く設定する。
- (6) サーバはユーザから各項目とそれらのチャート値を受信すると、「各項目に入力されたチャート値を要素とするベクトル」と「各ページの各センチメント値を要素とするベクトル」との  $\cos$  相関値を算出する。この  $\cos$  相関値を基にリランキングした結果をユーザへ提示する(図3)。

例えば、ユーザが検索キーワード「東京都知事」を問い合わせ

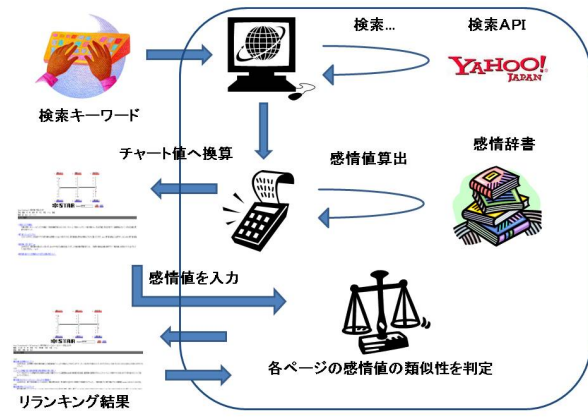


図 1 システムの概要

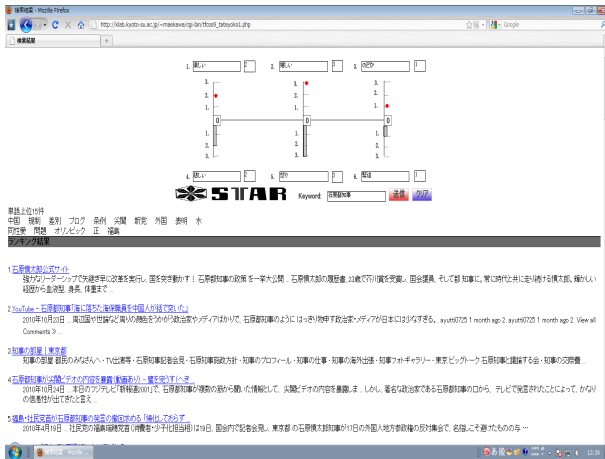


図 2 検索結果の例



図 3 ユーザのセンチメント値変更に伴うリランキング結果の例

せた結果、「楽しい」「嬉しい」「のどか」の各項目に対して、感情値の平均各々(2, 3, 1)であったとする。この時、ユーザが「楽しい」「嬉しい」「のどか」のセンチメントと対極のセンチメント「悲しい」「怒り」「緊迫」でリランキングしたい場合、該当項目の感情値を「7」「7」「7」へ変更する。それらの値を基に「東京都知事」の検索結果から「怒り(7)」「悲しい(7)」、「緊迫(7)」を基準としたリランキング結果を得ることができ

る。「東京都知事」に対して、ネガティブなセンチメントのページが上位にリランキングされる。

次に、システムを中心となるセンチメント値の算出(ステップ(3))と、センチメントに基づくリランキング方法(ステップ(6))を説明する。

### 2.2 センチメント値の算出

センチメント値の算出には感情辞書を使用する。感情辞書の作成方法を説明すると、まず感情軸を構成する感情語群を設定する(表 1)。次に、感情語を含む記事を抽出し、記事に含まれる感情語群  $IWL$  に属する感情語と感情語群  $IWR$  に属する感情語の数を比較し、 $IWL$  の数が多い記事の集合を  $S_L$ (記事数を  $N_L$ )、 $IWR$  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  $S_L$  における出現頻度を  $N_L(w)$ 、記事集合  $S_R$  における出現頻度を  $N_R(w)$  とすると、それぞれの補正済み条件付確率は、

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

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

と表される。この  $P_L(w)$  と  $P_R(w)$  を用いて、単語  $w$  のセンチメント  $s(w)$  を次のような式で表す。

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

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

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

表 1 感情語群

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

表 2 感情辞書一例

単語	感情軸 1	感情軸 2	感情軸 3
1 0	楽しい 悲しい	うれしい 怒り	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

センチメント  $s(w)$  は 0~1 の値をとる。1 に近い値は「楽し

い、うれしい、のどか」という感情を表し、0に近い値は「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」という感情軸のセンチメントは0.862であり、「楽しい」感情を表す。「偽装」という単語の「うれしい 怒り」という感情軸のセンチメントは0.075であり、「怒り」感情を表す。

ページのセンチメント値は、ページに出現した各単語のセンチメントの平均で算出され、0~1の値をとる算出値は、「楽しい」「嬉しい」「のどか」が強いと1に近づき、「悲しい」「怒り」「緊迫」が強いと0に近づくように設計されているが、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  
 という式を用いて同じスケールになるよう算出値を換算した。なお、負数の場合は絶対値をとり、ユーザに提示する。

### 2.3 センチメントに基づくランキング

システムは、ユーザから3つの項目を受信すると、各項目のチャート値をベクトルの要素とするベクトル  $V_q=(v_{q1}, v_{q2}, v_{q3})$  を決定する。次に、検索結果から取得した各ページに対して、3軸のセンチメントをベクトルの要素とし、各軸におけるセンチメント値を要素値とするベクトル  $V_p=(v_{p1}, v_{p2}, v_{p3})$ 、( $p=1, \dots, N$ ,  $N$ は取得した総ページ数)を決定する。以上のベクトル  $V_q$  と  $V_p$  の相関を下記より算出する。

$$\text{sim}(V_q, V_p) = \frac{(v_{q1}v_{p1} + \dots + v_{q3}v_{p3})}{\sqrt{(v_{q1}^2 + \dots + v_{q3}^2) * (v_{p1}^2 + \dots + v_{p3}^2)}} \quad (6)$$

この  $\cos$  相関値の高い順にランキングした結果をユーザへ提示する。つまり、チャート値を最大値3にしたときは、そのセンチメントに対する値が最も高いページが上位にランキングされる。

## 3. 実 装

実装した提案システムについて検討する。提案システムは、PHP Version 5.2.6, FLASH ver8.0 で開発し、検索エンジンには、Yahoo!検索 WebAPI Ver.1.0 を用いた。検索結果数は先行研究より Yahoo!の検索結果から得られた約 2000 件のうちサンプリングした 86 件とした。

### 3.1 インタフェース

検索結果の例を図4および図5に示す。

ユーザが入力した検索キーワードから Yahoo!の検索 API を用いた検索結果を出力し、同時に感情辞書を用いて検索結果全体の平均値を算出値をフラッシュ上に表示する(図4)。次に図4のフラッシュ上でユーザが入力したセンチメント値に基づいて、結果がランキングされ入力したセンチメント値に近いセンチメント値を持った記事から順に並べ替えられる(図5)。実験より、センチメント値の抽出およびユーザのセンチメント値変更によるランキングを確認できた。

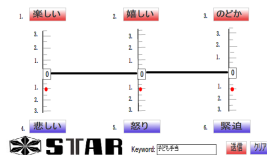


図4のフラッシュ上には、検索結果全体のセンチメント値の平均値が表示され、ユーザが入力したセンチメント値に近い結果が上位に並び替えられる。図5は、検索結果全体のセンチメント値の平均値を表示する。次にランキングが正しく行われているかを確認するため、その各々の平均値とは対極のセンチメント値を入力しランキングを行う。例えば「子供手当」を検索した際、検索結果全体から各軸の平均値(1軸=4.33192, 2軸=5.06413, 3軸=4.77202)が算出される。各軸の対極の値を取ることでランキングをする際に入力する値としては(1軸=-4, 2軸=1, 3軸=1)を選択する。同様に「裁判員制度」を検索した場合には平均値(1軸=4.6119, 2軸=4.81456, 3軸=4.63002)に対して(1軸=1, 2軸=1, 3軸=1)を入力してランキングを行った。

図4 検索結果全体とセンチメント値の平均値の例

### 3.2 実 験

本システムの適合性の評価を行った。実験では入力したセンチメント値に基づいたランキング結果の評価を行う。20代男性5人がランキング結果の各タイトルとスニペットを読んで、3つの感情軸に対して7段階評価を行った。その評価結果とシステムが算出した結果を図6図7に示す。

評価実験では例として、検索キーワード「子供手当」「裁判員制度」に対するランキング結果から評価と考察を行う。

評価の方法としては、まず例の検索キーワードを検索し、その際に算出される検索結果全体のセンチメント値の平均値を表示する。次にランキングが正しく行われているかを確認するため、その各々の平均値とは対極のセンチメント値を入力しランキングを行う。例えば「子供手当」を検索した際、検索結果全体から各軸の平均値(1軸=4.33192, 2軸=5.06413, 3軸=4.77202)が算出される。各軸の対極の値を取ることでランキングをする際に入力する値としては(1軸=-4, 2軸=1, 3軸=1)を選択する。同様に「裁判員制度」を検索した場合には平均値(1軸=4.6119, 2軸=4.81456, 3軸=4.63002)に対して(1軸=1, 2軸=1, 3軸=1)を入力してランキングを行った。

この結果を用いて被験者には、ランキングした結果上位10件の各タイトルとスニペットを提示し、各々3軸のセンチメント値において7段階で評価してもらった。被験者が評価した各軸ごとのセンチメント値から平均値を算出し、ランキングされた結果がランキングの際に入力したセンチメント値にどの程度近似して並べ替えられているか評価することによりランキングの適合性を判断できると考えた。(図6)

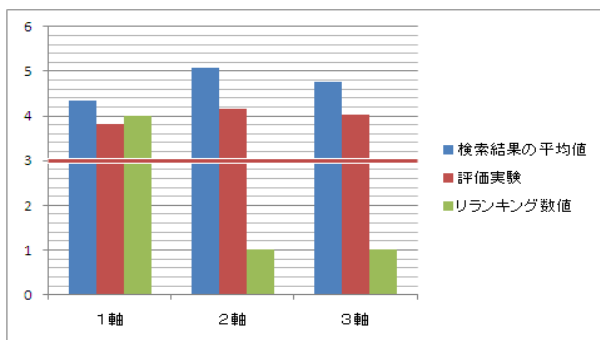


図6 検索キーワード「子供手当」におけるランキング結果

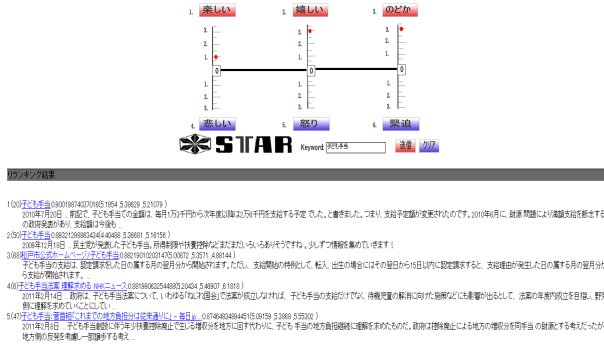


図 5 ユーザのセンチメント値変更に伴うリランキング結果の例

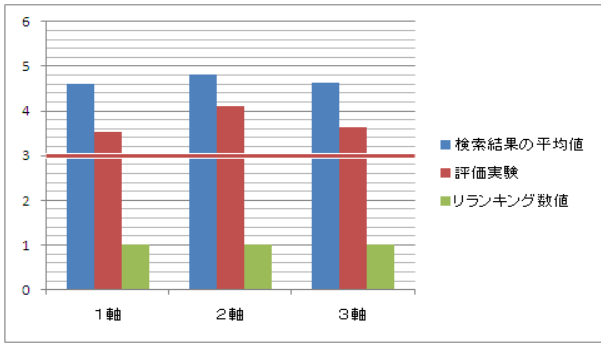


図 7 検索キーワード「裁判員制度」におけるリランキング結果

- C7-2, (2007)
- [7] レーダーチャート検索システム  
<http://klab.kyoto-su.ac.jp/%7Enogawa/>
  - [8] 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

以上の評価実験より, リランキングの結果としては図 6 図 7 より有効性が見られた. 今回はリランキング後の出力を可能な限り明確にするため, リランキングの際に入力した値を対極となる数値を最大値に定めたことが有意な差異となったのではな

#### 4. ま と め

本研究では, 検索キーワードに対するセンチメント値を提示し, それらの各値をインタラクティブに修正することでセンチメントに基づくリランキングを実現する検索システムを提案した. 評価実験の結果より, 感情値を用いた本システムのリランキングシステムには有効性があった. 本手法により, 検索結果の表層のテキストデータには含まれない深層部分のセンチメント値の算出や対極性を利用したリランキングが可能となった.

#### 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, 電子情報通信学会第 18 回データ工学ワークショップ (DEWS2007),