

# ニュース映像のシーン順序に基づく Blog 検索方式

北山大輔<sup>†</sup> 角谷和俊<sup>††</sup>

現在、ニューストピックについての情報はテレビ番組や Web サイトのニュース映像から得ることができる。ニュース映像にはいくつかのシーンが含まれ、その構成要素となっている。ユーザはこれらのシーンに含まれる映像や構成に基づき、そのニュースの内容を理解することが可能である。しかしながら、ニュース内で編集されたシーンの順序によって、ユーザの評価が変化する場合がある。すなわち、1つのニュースに対して複数の評価が存在することが考えられる。このような複数の評価に対し、対応する Blog 記事を検索し提示することで、映像のみでは得られない、ニュースの評価に関する情報が得られる。そこで我々は、シーンの出現パターンにより評価の抽出を行い、その評価に対応する Blog 記事を検索する方式を提案する。本稿では、シーン順序に基づくニュースの評価手法、および評価に基づく Blog 検索手法について述べる。また、プロトタイプシステムによる評価実験を行うことで、本手法の有効性を示す。

## A Blog Search Method Using News Video Scene Order

DAISUKE KITAYAMA<sup>†</sup> and KAZUTOSHI SUMIYA<sup>††</sup>

We often view news programs on TV or news video streams on Web sites to obtain information on current news topics. A video stream generally contains several scenes, i.e., a scene is a constituent element, and a viewer can gain a broader understanding of the topic by viewing all the scenes. Since each scene may provide a different viewpoint about the topic, the order in which they are viewed can affect the viewer's opinion of the topic. When a news video stream does contain various opinions, it may be possible to retrieve and read Blog entries corresponding to them. Because the information presented in the stream may be insufficient for a viewer to fully evaluate the opinions expressed, we have developed a method for finding Blog entries expressing particular points of view regarding the topic. It uses a form of evaluation and categorization similar to that based on the news scene order. In this paper, we describe our method based on news scene order and Blog search using news evaluation. Furthermore we show its effectiveness of the method by examining a prototype system.

### 1. はじめに

近年、ニュース映像は TV だけでなく、CNN.com International<sup>1)</sup>、CBSNews.com<sup>2)</sup>、ABC News<sup>3)</sup>、ANN NEWS<sup>4)</sup>、FNN-NEWS.COM<sup>5)</sup> など様々なニュースサイトでも配信されている。複数のニュースサイト間では、同一トピックのニュースであっても報道スタンスが異なる場合がある。また、ニュース映像の構成の差異により受ける印象が異なる場合がある。いい換えれば、意見の対立や異なる立場など複数評価を持つニュース映像では、シーンの構成順序によって、

映像から受ける評価の印象が変化する場合があると考えられる<sup>6)</sup>。たとえば、「鯨肉の給食が再開される」という捕鯨問題に関するニュースがあるとする。前半シーンで、「おいしい」「日本文化として喜ばしい」のように歓迎する意見が述べられていても、後半シーンに「今の日本に鯨を食べる習慣はない」という否定的な内容の意見があれば、あまり歓迎されていない印象を受けると考えられる。反対に、前半で否定されていても後半で歓迎の意見があれば歓迎されている印象を受けると考えられる。また、同一トピックに対して同じ映像素材を用いたニュース映像であっても、報道スタンスによりその順序構成が異なる場合がある。このようにシーン構成順序が異なると、ユーザが受ける評価の印象に差異が生じると考えられる。

一方、Weblog ( Blog ) の中には時事ニュースに対して意見を述べたり、評価を行ったりしているものも少なくない。評価が分かれるようなものは、Blog で

<sup>†</sup> 兵庫県立大学大学院環境人間学研究所  
Graduate School of Human Science and Environment,  
University of Hyogo

<sup>††</sup> 兵庫県立大学環境人間学部  
School of Human Science and Environment, University  
of Hyogo

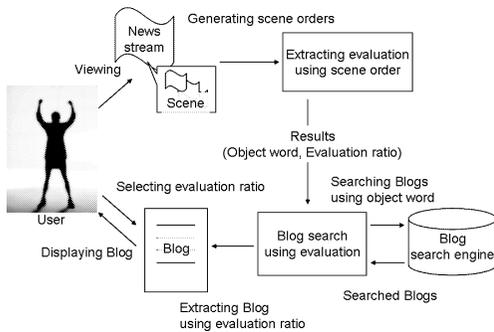


図 1 評価抽出に基づく Blog 検索の概念図

Fig. 1 Concept of Blog search using evaluation.

議論が交わされることがある。また、同じトピックであっても、1つのBlogで交わされる議論の肯定・否定の割合が異なる場合も存在する。

ニュース映像では、時間的な制約から、多面的な評価など、十分な解説がなされないことが多い。また、1つのニュース映像だけでは、異なる報道のスタンスを視聴することができない。同一トピックに対する複数のニュース映像を視聴することも考えられるが、冗長なシーンも多く、現実的ではない。そこで、ニュース映像のシーン順序によりニュースの評価の割合を抽出し、その評価割合に合致するBlogを検索する手法を提案する<sup>7)</sup>。

図1は、本手法の概念図である。まず、ユーザが視聴しているニュース映像に対しシーン分割を行う。次に、シーン順序を任意に入れ替えた仮想的なニュース映像を生成する。さらに、生成されたニュース映像の評価割合を算出する。最後に、得られた評価を用いてその評価に適合するBlogの検索を行い、ユーザに検索結果を提示する。その際、ユーザは単一のニュース映像から得られたキーワードと評価の度合いを選択することで、同一トピックでも異なった評価のBlogを閲覧することができる。このように、ニュース映像から得られた評価割合に合うBlog記事を得ることで、単一の映像ニュースのみでは不足している様々な角度からの評価を得ることができる。

以下に本稿の構成を示す。まず、2章では本研究の概要と関連研究について説明する。3章ではニュースの評価に基づくBlog検索について説明する。4章では作成したプロトタイプシステムについて説明し、評価実験とその考察を行う。最後に5章でまとめと今後の課題について述べる。

## 2. 本研究の概要と関連研究

### 2.1 本研究の概要

ニュース映像は、配信される前に作成者によりシーンの構成が決められる。我々は、シーンの順序によってユーザの受ける評価の印象が異なる場合が多いと考えている。たとえば、映画ではモンタージュ理論のクレショフ効果という現象がある<sup>(6),9)</sup>。この効果は、ユーザの受ける印象が、提示されるショットの順序によって変化することを示している。たとえば、男と棺と食べ物の3つのショットがあるとすると、食べ物のショットの後に男のショットがくれば、空腹という印象を感じると考えられ、一方、棺のショットの後に男のショットがくれば、哀れみという印象を感じると考えられる。いくつかのシーンで構成されるニュース映像でも同様の作用があると考えられる。また、Blogにおいてもコメントなどによる編集の結果で、ユーザの受ける評価の印象が変化すると考えられる。本手法の特徴は以下のとおりである。

- ニュースのシーン順序の自動生成

ニュースのシーン順序により、評価が変化するようなシーン順序を自動的に生成する。また、評価に影響を与えやすいシーンを算出し、算出された値に基づき順序を生成する。このことにより、仮想的にそのニュース映像から得られる可能性のある評価を算出することが可能になる。

- 対象語と評価表現の抽出

評価の構成要素として、“対象語”と“評価表現”の抽出を行う。生成された順序ごとに対象語・評価表現の重要度を算出する。さらに、評価表現に関しては肯定と否定の割合を計算し、評価割合を算出する。

- ニュースの評価に基づくBlogの検索

ニュース映像より抽出された対象語と評価割合を用いて、Blogの検索を行う。対象語による検索の後、Blog中の評価表現の重要度算出を行い、ニュースから算出したそれぞれの評価割合に合うBlogを、検索結果として提示する。

本手法により、ニュース中のトピックについて評価を行っているBlogを、ニュース中の評価割合に基づき検索することが可能になる。すなわち、ユーザは1つのニュースを視聴するだけで、そのトピックについての様々な評価をBlogから得ることができる。

モンタージュが有する“新しい意味を創造する働き”を特に強調して“クレショフ効果”と呼ぶことがある<sup>8)</sup>。

本稿では、これ以降「評価の印象」を単に「評価」と記する。

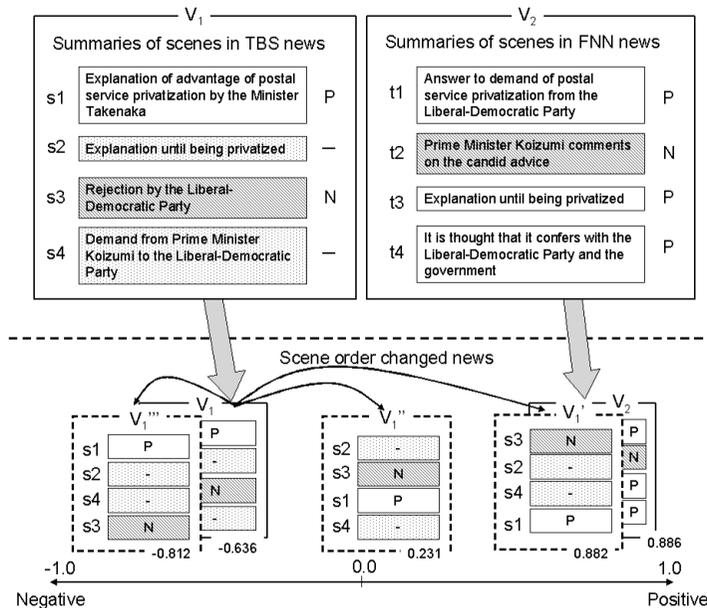


図 2 予備実験 1  
Fig. 2 An example of preliminary experiment.

## 2.2 予備実験

### 2.2.1 予備実験 1

報道各局の同一トピックのニュースでは、たとえ同じニュース映像素材を用いた場合でもその評価が異なる場合が存在する。たとえば、郵政民営化のニュースについて、TBS<sup>10)</sup>では小泉総理が苦言を呈するシーンを最後に用い、自民党からは不満の声が出ているというように否定的なニュースを構成している。一方、FNNでは小泉総理の苦言シーンを前半部に用い、自民党に関しては内容を協議するという表現にとどめている。これらは同一トピックに対して、同じニュース映像素材を用いているが、構成順序が異なる例である。このとき、ユーザは異なる印象を持つと考えられる。また、逆に、シーン順序を変更するだけで、他方のニュースと同一の印象を持つ仮想的なニュースを生成できることになる。そこで我々は、TBSのニュースのシーン順序を入れ替えて、FNNと類似したシーン順序を作ることで、FNNに近い評価を生成することが可能であるかを確認する予備実験を行った。

予備実験のデータを図2に示す。図2上部の四角  $V_1, V_2$  はオリジナルのTBS, FNNのニュース映像である。その中の  $s_1, s_2, t_1, t_2$  などはそれぞれのシーンであり、それぞれの四角はシーンの要約である。右側にある“P”, “N”, “-”は、それぞれのシーンに含まれる評価表現より決定した“肯定”, “否定”, “評価なし”のシーンである。図2下部の実線の四角はTBS,

FNNのオリジナルのニュース映像を表し、破線の四角はTBSのシーン順序を入れ替えることによって生成した仮想的なニュース映像である。それぞれの四角の下部の数字は、算出した評価割合である。四角が左にあるほど否定的な評価となることを表し、右にあるほど肯定的な評価となることを表す。実験の結果、TBSより生成したいくつかの評価割合のうち  $V'_1$  の評価割合 (0.882) が FNN の  $V_2$  の評価割合 (0.886) と近くなることが確認された。

この予備実験により、我々はニュースのシーン順序を変えることで、評価を変化させることができると考えた。また、このことによりシーン順序を変えることで1つのニュースから、異なる報道局における同じトピックのニュースの評価を得ることができるようになる。すなわち、シーン順序を入れ替えることによって得られる評価は、他の報道局のニュースで述べられている評価と類似する場合があり、1つのニュース映像から、他局のニュースについて述べたBlogの検索が行えることを意味する。予備実験の他の例では、「靖国神社の合祀」に関する裁判のトピックや、「外務大臣の外交力」に関するトピックで、同じニュースの素材を用いているが、シーン構成の順序が異なる場合を確認した。

3章で提案する評価割合の算出方法で計算を行った。その結果  $V_1, V'_1, V''_1, V'''_1, V_2$  それぞれの値が  $-0.636, 0.882, 0.231, -0.812, 0.886$  となった。

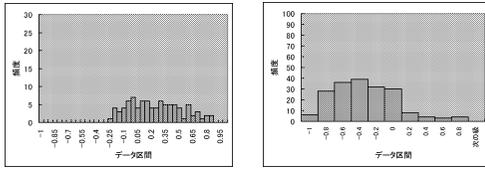


図 3 評価割合 (自民党総裁選)

Fig. 3 Histogram of evaluation ratio.

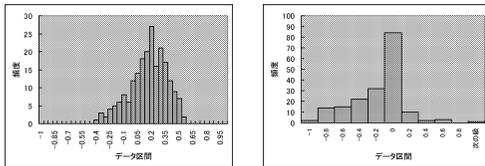


図 4 評価割合 (貸金業の規制改正案)

Fig. 4 Histogram of evaluation ratio.

他の報道局の評価が得られるかを確かめるため、4つの報道局 (TBS・FNN・ANN・NNN<sup>11)</sup>) が報じた「外務大臣の外交力」に関するトピックのニュースを用いて、1つのニュースから生成された複数の評価割合が他の報道局をカバーすることが可能であるかという予備実験を行った。実験は、TBSのニュースのシーン順序を自動的に生成したものと、他の局のものとのシーン順序の評価割合を比較することで行った。その結果、TBSのニュースから生成された評価割合は1.0, 0.4, 0.2, 0.0, -0.2, -0.4, -0.6の値であり、各局の評価割合はANNが0.2, NNNが-0.2, FNNが0.4となった。これらの各局の評価割合は、TBSから生成された仮想的な評価割合に含まれていることが確認できた。これらのことより、シーン順序の自動生成の意義は、1つのニュース映像からでも、評価の異なる他局のニュースを仮想的に生成し、その評価に合致するBlogを検索可能であることであると考えられる。

### 2.2.2 予備実験 2

1つのニュース映像からシーンの組合せにより自動的に生成された評価割合が、同一トピックのBlogの評価割合と一致するかどうかの予備実験を行った。その結果、大きく2つのパターンが見られた。一方は、ある評価点を中心に件数が増えるパターン、他方は、平坦な分布となるパターンである。ニュースから生成された評価割合と、ニュースのトピックについて述べているBlogの評価割合の度数分布を図3、図4に示した。図の左側がニュース映像のシーン順序によって生成された評価割合のヒストグラムであり、右側が

そのトピックに関するBlogの評価割合のヒストグラムである。図3は平坦な分布となるパターンの例であり、図4はある評価点を中心に件数が増えるパターンの例である。図のように、Blogの分布に対応してニュースの評価割合が生成されていることが分かる。これらの結果によりシーン順序を自動的に生成することで、適切な評価を持ったBlogの検索を行うことが可能であると考えられる。

### 2.3 関連研究

Maら<sup>12)</sup>やHenzingerら<sup>13)</sup>は映像の音声テキストを用いて自動で検索質問を作成し、Webより情報を取得する手法を提案している。音声テキストから検索質問のキーワードを抽出し、検索を行うという点で本研究と類似しているが、ニュース映像中の評価に基づき、Blogの検索を行うという点で目的が異なる。

Nannoら<sup>14)</sup>はblogWatcherというBlog検索に特化したシステムを構築している。blogWatcherでは、Blog検索以外に、キーワードのburst度、ホットキーワードの表示、評判情報検索の機能を有している。このうち、評判情報検索が入力キーワードに関する評判情報をBlogから検索して提示するという部分が、本研究のBlog検索と類似している。しかし、Blog中における議論の肯定・否定の割合に着目し、評価割合の一致により検索を行っている点で本研究とは異なる。

Fukuharaら<sup>15)</sup>はBlog記事を分析することで人々の関心の傾向の抽出を行っている。実世界における現象とBlog記事数を用いて分析し、関心を示すトピックの抽出を行い、傾向を可視化してユーザに提示する。Glanceら<sup>16)</sup>は収集したBlog記事からトレンド抽出を行うBlogPulseというシステムを提案している。一定期間におけるトピックを表すキーワード、主要人物を表すキーワードの出現頻度から、そのときにトレンドとなっているかどうかを判別している。これらの研究では、Blog記事の傾向の分析を行い、結果を提示するが、本研究ではBlog記事で行われている評価によって、ニュース映像のトピックの評価を補うことを目的としている点で異なる。

Kumuarら<sup>17)</sup>はハイパーリンクによるBlog群のつながりをblogspaceと定義し、Blogコミュニティを抽出し、コミュニティの進化に関する研究を行っている。Gruhlら<sup>18)</sup>はマクロな視点、ミクロな視点からのトピック伝達の特徴づけを行い、議論が交わされているBlogと外部的要因により発生するBlogによりBlogにおける情報伝播のモデル化を行っている。ただし、これらの研究は、情報の伝播に着目したものであり、Blog中のトピックに対する評価を扱うもので

Blogの評価割合に関しては、被験者が0.2刻みで判定した。

はない。

評価表現の抽出および評価判別に関する研究として、Dave ら<sup>19)</sup>、立石ら<sup>20)</sup>の研究があげられる。Dave らは Web 上のレビューを肯定・否定に分類し、抽出する手法を提案している。この手法では、レビューサイトをコーパスとすることで、肯定・否定の指標としている。本研究では、対象をニュース上のトピックとしているため、製品情報などの評判情報とは表現が異なると考えられる。ただし、評価表現の辞書の作成に関しては、キーワードベースで評価の判別を行う立石らのアプローチを参考にしている。

### 3. ニュースの評価に基づく Blog 検索

#### 3.1 シーン分割と評価表現

本手法ではシーンの順序生成を行うために、まず、シーンの分割を行う。本手法では従来の画像処理や音声認識などを組み合わせて分割を行う<sup>21),22)</sup>。まず、音声の無音区間と、映像の色特徴の変化点を抽出する。近傍に映像の色特徴変化点のある音声の無音区間をニュース映像のシーンの分割点とし、分割を行う。

評価表現を判別するために、評価表現辞書を作成する<sup>20)</sup>。辞書には、“歓迎”、“批判”のような肯定もしくは否定の意味を含む単語を肯定・否定という情報とともに登録している。また、株価の“上昇”、“下落”など、一見事実であるように見えるが、評価ともとれるものは評価表現として辞書に登録している。映像の音声テキスト中に出現する単語すべてに対し、評価表現辞書に登録されている単語か否かを判別し、登録されている単語に関して評価の肯定・否定の判別を行った。

また、評価の補助表現として“ない”のように評価表現と組み合わせることで評価が反転するような単語も、補助表現として登録している。評価の反転に関しては、補助表現より前 4 単語中最も近い評価表現の肯定・否定を反転させる。4 単語以内に評価表現が存在しない場合、その補助表現は評価に影響を与えない。なお、疑問文中に出現する評価表現には評価の意図はない場合があると考えられるが、今回は考慮していない。なお、映像の音声テキストに関しては、クローズドキャプションなどにより与えられているものとしている。

#### 3.2 組合せ数の削減

シーン順序生成を自動で行う場合、すべてのシーンについて総当たりで行うと組合せ数が多くなり、計算量が増加するという問題がある。シーン数の多い映像では、膨大な計算が必要となるため、シーン操作の組合せ数を適切に削減する方法が必要となる。そこで、

シーンの評価に基づいて、シーンの重要度を算出し、重要なシーンのみの組合せを生成することにより、組合せ数の削減を行う。具体的には、 $i$  番目のシーン  $s_i$  について、シーン中に含まれる  $j$  番目の評価表現  $e_{ij}$  をシーン内の出現順により重要度算出を行う。算出された評価表現の重要度を合計することによって、シーン評価値  $Scene\_eval(s_i)$  を決定する。

$$Scene\_eval(s_i) = \sum_{j=1}^{E_i} eval(e_{ij}) \quad (1)$$

$$eval(e_{ij}) = \alpha \times \exp(-(E_i - j + 1)) \quad (2)$$

$$\alpha = \begin{cases} 1.0 & \text{if } e_{ij} \text{ is positive word} \\ -1.0 & \text{if } e_{ij} \text{ is negative word} \end{cases} \quad (3)$$

式中の  $E_i$  はシーン  $s_i$  内の評価表現出現数であり、 $\alpha$  は評価表現の肯定・否定を表す重みである。 $Scene\_eval(s_i)$  が正の値をとればシーン評価を肯定、負の値をとれば否定とする。評価表現の含まれないシーンの場合  $Scene\_eval(s_i)$  がゼロとなり、この場合はシーン評価を評価なしとする。

手順としては、まず、シーン評価の否定、肯定、評価なしの 3 要素からなるすべての組合せを作成する。その後、シーン評価が同じものに関して、評価値  $Scene\_eval(s_i)$  の絶対値が大きいものほど後のシーンとすることで、組合せの削減を行う。評価値が大きいシーンを後ろとするのは、ニュース全体としての評価に影響を与えやすくするためである。すなわち、評価に影響を与えないような組合せの削減を行っている。評価なしのシーンに関しては、順序は基となるニュース映像のシーン順序とした。総当たりであればシーン数の階乗個生成される組合せを、この手法によりシーン数の階乗を肯定シーンの階乗、否定シーンの階乗、評価なしシーンの階乗の 3 つを乗算した値で除算した数の組合せに削減することができる。

肯定シーン  $P$  が 2 つ、否定シーン  $N$  が 2 つ、評価なしのシーン  $-$  が 1 つの場合の例を図 5 に示した。各シーンの評価値の大小関係は  $P_1 > P_2$ ,  $N_1 > N_2$  である。この例では、総当たりの組合せが 120 組生成されるのに対し、30 組に削減することが可能である。

#### 3.3 シーン順序に基づく評価抽出

##### 3.3.1 対象語の抽出

評価は対象語と評価割合により構成される。対象語とは、評価の対象となるトピックを表す名詞である。ただし、名詞のうち対象となりえない品詞、たとえば代名詞などは除く。ニュース映像においては、シーン順序が先に出現するほど重要であり、シーン内でも先

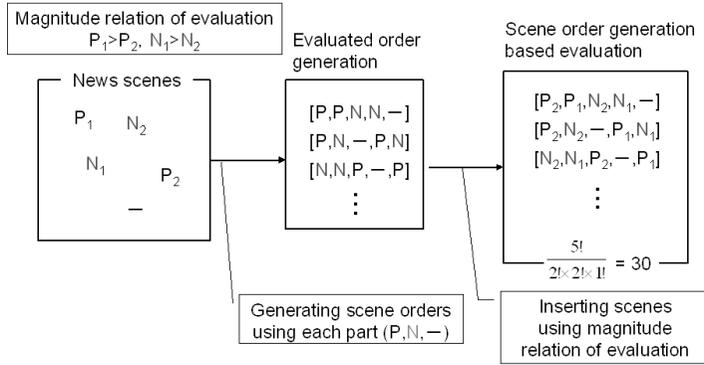


図 5 評価を考慮した組合せの削減  
Fig. 5 Reduction of combinations.

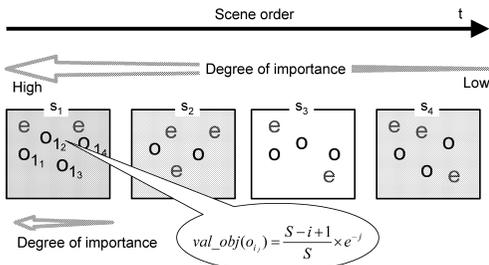


図 6 対象語の重要度の算出  
Fig. 6 Degree of object word importance.

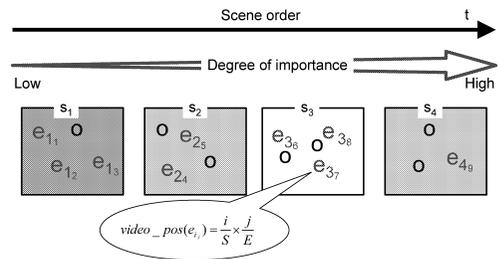


図 7 評価表現の重要度の算出  
Fig. 7 Degree of evaluated expression importance.

に出現するキーワードが対象語として重要であるといえる。また、ニュース内で複数回出現するキーワードも、対象語として重要といえる。そのため、シーン順序が先であるほど重要さが上がり、シーン内での出現順が先であるほど重要さが上がる関数を用いて抽出を行う(図6)。

シーン中の名詞  $o_{ij}$  について、シーン順重要度とキーワード出現順重要度を乗算し、重要度  $val\_obj(o_{ij})$  を算出する。同一キーワードに関しては、同一キーワードどうしの値を足し合わせたものを対象語の重要度とする。

$$val\_obj(o_{ij}) = \frac{S-i+1}{S} \times \exp(-j) \quad (4)$$

式中の  $S$  はシーン数、 $o_{ij}$  は  $i$  番目のシーンで  $j$  番目に出現する名詞である。名詞に関しては茶釜<sup>23)</sup> による形態素解析を用いて抽出を行う。

3.3.2 評価割合の算出

評価割合とは、評価表現の肯定・否定重要度の割合である。手順としては、まず、評価表現の重要度の算出を行い、その後評価割合を算出する。評価表現は対象語とは異なり、ニュース映像内での出現順序が後であるほど、ユーザに対して強い印象を与えると考えられる。たとえば、「環境保護団体は鯨保護のために捕鯨

に反対する」と「環境保護団体が捕鯨に反対なのは鯨保護の精神からである」であれば、前者のほうが反対という印象が強いといえる。そのため、シーン順序が後であるほど評価の重要さが上がり、かつ、ニュース映像内のすべてのキーワードの中での評価表現の出現順序が後であるほど重要さが上がる関数を用いて重要度の算出を行う(図7)。

ニュース映像中の評価表現  $e_{ij}$  について、シーン順序とニュース映像内出現順序を用いて重要度  $video\_pos(e_{ij})$  および  $video\_neg(e_{ij})$  を算出する。評価表現は対象語とは異なり、同一キーワードでも評価が反転する場合が考えられるため、同一キーワードの足し合わせは行わない。

$$video\_pos(e_{ij}) = \frac{i}{S} \times \frac{j}{E} \quad \text{if } e_{ij} \text{ is positive word} \quad (5)$$

$$video\_neg(e_{ij}) = -1.0 \times \frac{i}{S} \times \frac{j}{E} \quad \text{if } e_{ij} \text{ is negative word} \quad (6)$$

式中の  $S$  はシーンの総数、 $E$  はニュース映像内の評価表現の総数、 $e_{ij}$  は  $i$  番目のシーンに出現する  $j$  番目の評価表現である。

評価表現は評価表現辞書を用いて抽出を行う。評価表現の重要度とは、そのニュース映像内でユーザに

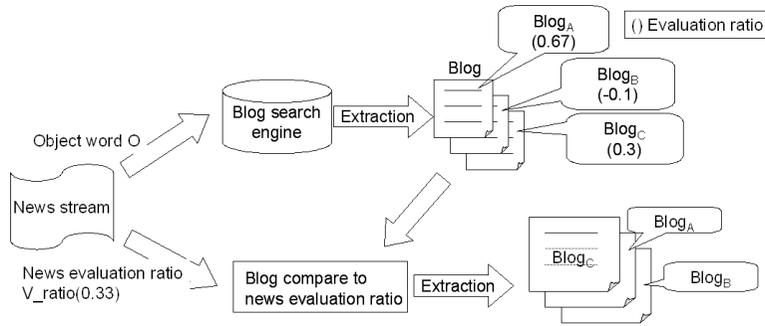


図 8 評価に基づく Blog 検索  
Fig. 8 Blog search using evaluation ratio.

与える印象を表す値で、大きければニュース映像内でユーザにその評価の印象を与えている可能性が高い。出現するすべての評価表現中、肯定表現の重要度を足し合わせた値を  $P_{video}$ 、同じく否定表現の重要度を足し合わせた値を  $N_{video}$  とし、以下の式でシーン順序の評価割合  $V\_ratio$  を算出する。

$$V\_ratio = \frac{P_{video} + N_{video}}{|P_{video}| + |N_{video}|} \quad (7)$$

$$P_{video} = \sum_{i=1}^S \sum_{j=1}^E video\_pos(e_{ij}) \quad (8)$$

$$N_{video} = \sum_{i=1}^S \sum_{j=1}^E video\_neg(e_{ij}) \quad (9)$$

式中の  $S$  はシーンの総数であり、 $E$  はニュースに含まれる評価表現の総数である。評価割合のとりうる値は  $-1.0$  から  $1.0$  の間であり、 $-1.0$  に近いほど否定寄り、 $1.0$  に近いほど肯定寄りなシーン順序の組合せとする。

### 3.4 評価を用いた Blog 検索

#### 3.4.1 Blog 中の対象語重要度

ニュース映像の対象に対する評価を検索する場合、一般の検索エンジンでは、検索キーワードが出現する Web ページを検索対象とするため、個人の行った評価ではない Web ページが検索されてしまう。また、従来の Blog 検索エンジン、たとえば Yahoo!Blog<sup>24)</sup> や gooBlog<sup>25)</sup> では、検索対象を Blog に限定することができるが、検索キーワードが Blog のエントリー中に含まれている Blog を抽出して提示のみにとどまる。すなわち、ニュース映像の対象とその評価を議論する Blog を検索するには、キーワードベースの手法では限界があると考えられる。ユーザが肯定的な単語や否定的な単語を入力して検索することも考えられるが、ユーザの入力できる単語の種類には限界があり、また、その単語が Blog 中に含まれているとは限らない。

評価を用いた Blog 検索の目的とするところは、ニュース映像から抽出されたトピックについて、そのニュースから生成される評価と同等の評価を持つ Blog を取得することにある。たとえば、捕鯨のニュースに関して、鯨肉の給食について否定寄りの印象を受けるシーン順であれば、取得される Blog はエントリーで鯨肉給食について扱っており、かつコメント・トラックバックを含めた議論が否定寄りで展開されていることが望ましい。

そこで本手法では、評価を用いた Blog 検索を実現するために、従来の Blog 検索エンジンによる検索結果を用いて、対象語に基づきニュースのトピックに合致する Blog の抽出を行う。次に Blog の評価割合を算出し、ニュースの評価割合にあった Blog の検索を行う。図 8 は、評価を用いた Blog 検索の手順を图示したものである。

まず初めに、Blog に含まれる対象語の重要度について述べる。重要度の計算には Blog の構成のうち、エントリータイトル、エントリーを用いる。また、ニュースについて意見を述べるような Blog の場合、エントリー中でニュースそのものに対して引用の代わりにリンクを行っている場合（以降、引用リンクと呼ぶ）があり、引用リンクとエントリー本文は区別して扱う。エントリータイトルは、トピックを表すようにつけられている場合と、感想を述べるようにつけられている場合がある。トピックを表すようにつけられている場合は、エントリー中にも同じキーワードが出現しやすく、感想を述べているような場合は、エントリー中に出現しにくいという傾向が見られる。また、引用リンクに用いられる文字列は、リンク先のニュースのタイトルを用いることが多く、トピックを表しているといえる。これらのことより、タイトルに含まれるか、引用リンクに含まれるかということを手がかりに対象語の重要度の算出を行う。

$$\text{blog\_obj}(o_i) = w \times \text{count}(o_i) \quad (10)$$

$$w = \begin{cases} \beta & o_i \in (T \cap A) \\ \gamma & o_i \in ((T \cap E) \cup A) \\ \delta & o_i \in E \end{cases} \quad (11)$$

式中の  $o_i$  は Blog 中に含まれる  $i$  種類目の名詞であり、 $\text{count}(o_i)$  はその出現回数である。  $w$  は出現位置による重みであり、 $o_i$  の出現位置により  $\beta > \gamma > \delta$  の値をとる。  $T, A, E$  はそれぞれ、エントリのタイトル、引用リンク、エントリ本文の単語の集合である。

### 3.4.2 Blog 中の評価表現重要度

次に、Blog における評価表現の重要度について述べる。 Blog では、ニュースとは異なり、書き手により様々な書き方が可能である。 そのため、Blog 全体としてどのような評価を述べているのかを算出するには、Blog 中の出現位置を考慮したり、対象となっているキーワード付近を重要と見なしたりするのではなく、書き手本人の意見であるエントリと、それに対して議論するコメントの部分すべてを考慮する必要があると考えられる。 そのため、エントリおよび、コメントに出現する評価表現の個数を肯定、否定の重要度として算出する。

$$P_{\text{blog}} = \text{count}(e^e) + \text{count}(e^c) \quad \text{if } e \text{ is positive word} \quad (12)$$

$$N_{\text{blog}} = -1.0 \times (\text{count}(e^e) + \text{count}(e^c)) \quad \text{if } e \text{ is negative word} \quad (13)$$

$$e^e \in E \quad (14)$$

$$e^c \in C \quad (15)$$

式中の  $e^e, e^c$  は評価表現であり、 $E, C$  はそれぞれエントリ本文、コメントの単語集合である。  $\text{count}$  関数によって、それぞれに出現する評価表現の個数を算出している。 なお、個数以外の手法として、評価表現の出現の絶対位置により重要度を求めて算出する手法と、対象語と評価表現の相対位置により重要度を求める手法の比較を予備実験により行ったところ、個数により求める手法が他の手法よりも高くなった。

重要度の求め方を決定するために予備実験を行った。 比較した手法は以下の 3 つの手法である。

- 個数により求める手法：Blog 中の評価表現の個数を重要度としたもの
  - 絶対位置により求める手法：Blog 記事中で最後に書かれる評価表現ほど重要としたもの
  - 相対位置により求める手法：Blog 記事中に出現する対象語近傍の評価表現を重要としたもの
- 絶対位置により求める手法の評価表現の重要度は以下の式で算出した。

$$P_{\text{blog}} = \sum_{i=1}^n \exp(-(B - i + 1)) \quad \text{if } e_i \text{ is positive word} \quad (16)$$

$$N_{\text{blog}} = \sum_{i=1}^n -1.0 \times \exp(-(B - i + 1)) \quad \text{if } e_i \text{ is negative word} \quad (17)$$

式中の  $i$  は、Blog のエントリ、コメント、トラックバック中における評価表現の出現順序である。  $n$  は、Blog 中の評価表現の個数、 $B$  は、算出対象となるエントリ、個々のコメントに含まれる評価表現の個数である。 また、相対位置により求める手法の評価表現の重要度は以下の式で算出した。

$$P_{\text{blog}} = \sum_{i=1}^n \sum_{j=1}^m \exp(-\log(\text{dist}(e_i, o_j))) \quad \text{if } e_i \text{ is positive word} \quad (18)$$

$$N_{\text{blog}} = \sum_{i=1}^n \sum_{j=1}^m (-1.0 \times \exp(-\log(\text{dist}(e_i, o_j)))) \quad \text{if } e_i \text{ is negative word} \quad (19)$$

式中の  $e_i$  は Blog 中の全単語のうち  $i$  番目に出現する評価表現であり、 $o_j$  は Blog 中の全単語のうち  $j$  番目に出現する対象語である。  $n$  は、Blog 中の評価表現の個数、 $m$  は、Blog 中の単語の個数である。  $\text{dist}$  関数により、2 つの単語間の単語距離を算出している。

本実験の被験者は 5 人であり、400 件の Blog を用いて行った。 被験者によって正解のつけ方にばらつきがあるため、被験者による正解と各手法による解の一致数を被験者が判定した Blog 件数で除算して適合率を算出する。 評価の境界点となる値は以下のようにして求めた。

- (1) 被験者が各 Blog データの評価を判定し、評価の構成比を算出する。
- (2) 被験者が各 Blog データの評価割合を判定し、そのヒストグラムを作成する。
- (3) 得られたヒストグラムを、評価の構成比に基づいた面積比に分割をする。
- (4) その分割点となった評価割合を評価の境界点とする。

評価判定は肯定・中立・否定の 3 種類に分類した。 肯定・否定と判定した Blog はそれぞれトピックに対し肯定的・否定的なことを述べている Blog である。 中立と判定した Blog は、何らかの意見を述べているがどちらともいえないもの、エントリで肯定的に述べているが、コメントで反論されているような Blog である。 被験者は評価割合の判定は 1.0 から -1.0 の間で 0.2

刻みで判定した．1.0 に近いほど肯定的，-1.0 に近いほど否定的とし，Blog を閲覧して判定を行った．その結果，評価の構成比（肯定件数：中立件数：否定件数）は約 5:4:11 となり，肯定・中立の境界点として 0.36，中立・否定の境界点として -0.37 が得られた．この境界点を正解とし，各手法の比較を行った．実験の手順を以下に示す．

- (1) 被験者が各 Blog データの評価を判定し，それを正解とする．
- (2) 各 Blog データを各手法により肯定・中立・否定に判定する．
- (3) 各手法の判定結果と正解を比較し，適合率を算出する．

実験の結果，適合率は個数により求める手法が 45.3%，絶対位置により求める手法は 43.1%，相対位置により求める手法が 37.3% であった．そのため，本稿では個数により求める手法を用いる．

### 3.4.3 評価に基づく Blog 検索

映像より抽出された対象語と評価割合  $V\_ratio$  を用いて検索を行う．Blog 検索エンジンに対し，対象語を検索キーワードとして OR 条件で検索を行う．得られた Blog 集合に対し対象語の重要度計算を行い，検索に用いた対象語の重要度の総計を Blog のスコアとし，上位のものを対象 Blog とする．個別の対象 Blog 中のエントリ，コメントより評価表現を抽出し，重要度により評価表現の割合を算出する．Blog 中の肯定表現の重要度を足し合わせた値を  $P_{blog}$ ，否定表現の重要度を足し合わせた値を  $N_{blog}$  とし以下の式で Blog の評価割合  $B\_ratio$  を算出する．

$$B\_ratio = \frac{P_{blog} + N_{blog}}{|P_{blog}| + |N_{blog}|} \quad (20)$$

シーン順序の評価割合  $V\_ratio$  と対象 Blog の評価割合  $B\_ratio$  を比較し，割合の近いものを抽出し検索結果として提示する．これを，シーン順序により抽出される評価割合すべてに対して行う．この結果，あるニュース映像のシーンの組合せから自動的に生成されたいくつかの評価割合に基づき，対応する Blog が抽出される．

## 4. プロトタイプシステムと評価

### 4.1 システム構成

プロトタイプは Visual Studio.NET の C# を用いて実装し，動画の表示部分には Microsoft Media Player SDK を用いた．図 9 は実際のプロトタイプシステムの画面である．画面左上のニュースビデオ視聴部 (News video window) でニュース映像を視聴する



図 9 プロトタイプシステム

Fig. 9 Screen image of prototype system.

ことができる．対象語リスト (Lists of object word combinations) の中から抽出された対象語を選択することができ，評価割合バー (evaluation ratio) を操作することで，抽出された評価割合の選択を行うことができる．ニュースから生成された範囲の評価割合に関しては，数値の色を変化させることで選択する際に何を選擇できるのかということが分かるようにした．対象語リストは，キーワードの組合せであるため，よく似た組合せが出現する場合がある．そのため，対象語による重要度最上位の Blog のタイトルを併記することでユーザが選擇する際の手がかりになるようにした．Blog 閲覧部 (Retrieved Blog window) にはユーザの選擇に基づいた検索結果の Blog が提示される．また，システム内部は大きく分けてシーン順序生成部 (Scene order generation unit)，評価抽出部 (Calculated evaluation unit)，Blog 検索部 (Blog search unit) の 3 つからなる．図 10 にシステム構成を示す．

### 4.2 評価

#### 4.2.1 実験 1: ニュース映像の評価割合の精度

提案手法を評価するために，映像の評価割合算出の評価実験を行った．実験データは 2006 年 9 月 4 日～9 月 12 日の TBS, FNN のニュース映像 8 件 (表 1) を用い，被験者人数は 5 人である．映像の評価割合に関しては，被験者がニュース映像を視聴した後に，数値で評価割合を答える方式により行った．評価は正解率とし，正解件数を実験件数で除算して求めた．実験の手順を以下に記す．

- (1) 被験者がニュース映像を視聴する．
- (2) 被験者がニュース映像の評価割合を肯定の最大を 1.0，否定の最大を -1.0 とし，その間の数値で答える．
- (3) 映像の評価割合に関して被験者の解を正解とし，システムの算出した値と 20% の範囲内での一致

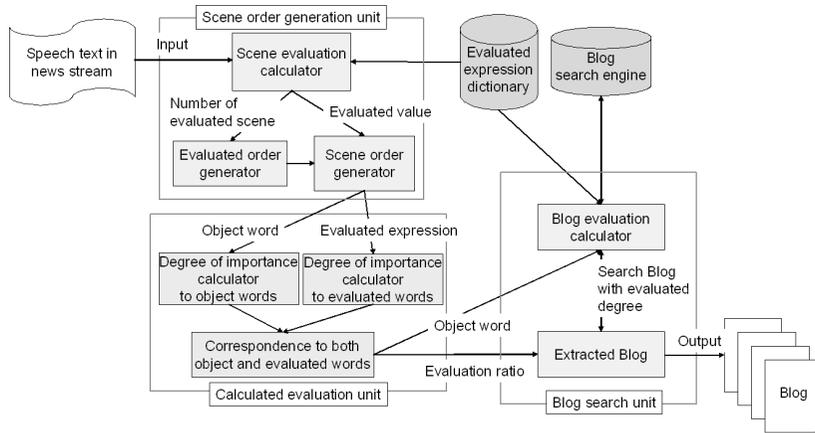


図 10 システム構成図

Fig.10 System architecture.

表 1 評価実験のニュースリスト

Table 1 News list.

ニュース番号	タイトル
1	自民党総裁選候補が公開討論会 安倍氏「参院選候補者を総裁選権限で差し替えも」
2	自民党総裁選 東京・秋葉原で安倍・谷垣・麻生の3候補そろっての初の街頭演説
3	「悪意に満ちた起訴状」 無罪主張し検察と全面対決
4	亡き夫の冷凍精子で出産 父子関係認めず
5	田中康夫流・・・ 知事 退任の一日
6	自民党、貸金業制度改革案で大詰めの協議 「特例金利」期間などで合意に至らず
7	自民党、金融調査会を再開し「特例金利」の期間短縮などをめぐり大詰めの議論へ
8	貸金業の規制改正案 自民党 来週に結論持ち越し

により正解率を算出する。

実験の結果を以下に述べる。映像の評価割合に関して、被験者ごとに大きな差は見られなかった。被験者間の評価の平均値とシステムが判定した評価とが一致した割合は、62.5%となった。一致したものの内訳は、5%の範囲内が2件、10%の範囲内が1件、20%の範囲内が2件である。不一致としたものの内訳は、30%の範囲内が1件、50%の範囲内が2件であった。判定が失敗するケースは、評価表現辞書に登録が困難である単語が含まれ、かつニュース全体への影響の大きい順序で出現している場合であった。たとえば、「日本一の尾道にすると」と期待していたが、結果がああいうことになって」と述べているようなシーンでは、シーンに含まれる評価は否定であると考えられるが、明示的に否定を示すようなキーワードが存在しない。このようなシーンが印象の評価に影響を与えやすい後半に出現している場合に、誤判定が多いと考えられる。誤判定を起こした場合、システムの算出する値は人間の感覚による値とは異なる場合がある。

これらから、本手法は評価を表すようなキーワードが映像内に含まれている場合、人間の感覚によるニュース映像の評価とほぼ近い値を算出することが可能であることが分かった。

4.2.2 実験2：ニュース映像から対象語抽出の精度

ニュース映像からの対象語抽出の精度を評価するため、提案手法と従来のキーワード抽出手法である単語の頻出度を用いる手法の比較実験を行った。各手法により検索した Blog から、被験者が評価の対象となっているキーワードの選択を行い、その適合率により評価を行う。ニュースは、表1のニュース番号1から5の5件で行い、Blogはニュースごとに各手法により得られたキーワードにより検索したものをを用いた、被験者人数は6人である。被験者は、検索に用いたキーワードが未知の状態で行った。実験の手順は以下のとおりである。

- (1) ニュース映像より、提案手法を用いて対象語、従来手法を用いて頻出語を抽出する。

被験者がつけた評価割合の相関係数の平均値は 0.60 であった。

対象語の語数は各ニュースより5種類程度抽出され、頻出語の単語数は対象語の単語数と同じになるように上位から抽出した。

表 2 ニュース映像から対象語抽出の精度

Table 2 Experimental result of object word extraction.

ニュース番号	従来手法 (頻出語)	提案手法 (対象語)
1	0.850	0.842
2	0.675	0.725
3	0.825	0.900
4	0.800	0.717
5	0.817	0.917

- (2) 各手法のキーワードを用いて Blog 検索を行い、その結果の上位 10 件ずつ、計 20 件を実験データとする。
- (3) 被験者が Blog を閲覧し、何を対象に評価を述べているかを判定し、キーワード 3 語で回答する。
- (4) 被験者の回答と各手法のキーワードが 1 語以上一致する場合を正解と見なし、実験データ数で除算することで適合率を算出する。

実験結果を表 2 に示す。多くのニュースで、従来手法とほぼ同じか、より良い結果が得られた。これは、従来手法で得られるキーワードは、それぞれのキーワードが共起することが多く、そのため、他の事柄を対象とした Blog でもよく使われるキーワードとなり、それらのキーワードが含まれている Blog であっても、対象として述べられていない場合が考えられるためである。提案手法で得られる対象語は、個々が独立して評価の対象を表すと考えらるため、それらのキーワードを含む Blog においても、評価の対象として述べられることが多い。このことにより、提案手法による対象語抽出は、ニュース中の評価の対象を表し、かつ検索される Blog の対象を表しているといえる。

#### 4.2.3 実験 3：ニュース映像と Blog の評価割合の一致に関する評価

提案手法のニュース映像と Blog の評価割合の一致による検索の評価を行うため、人手によりニュース映像と Blog を対応付けた際に相関が存在するかを確認する実験を行った。被験者は 4 人であり、実験に用いたトピックは「北朝鮮による拉致問題」、「村上ファンドの阪神買収問題」、「沖縄米軍基地編成問題」である。それぞれニュース映像は 5 件程度、Blog は 30 件程度で実験を行った。実験の手順は以下のとおりである。

- (1) 実験データとして同じトピックのニュース映像、Blog を手作業により収集する。
- (2) 被験者がニュース映像、Blog を視聴し、それぞれに評価割合をつける。
- (3) 被験者がそれぞれのニュースに対し、対応する Blog を選択する。

- (4) 対応付けされたニュースと Blog に対し、相関係数を求める。

- (5) 被験者に対し、評価割合をつける際の基準、対応付けを行う際の基準についてのヒアリング調査を行う。

実験結果として、人手により対応付けられたニュースと Blog の相関係数の平均値は 0.80 となり、相関が存在することが確認できた。各被験者の評価割合のつけ方に関しては、ニュース映像では各被験者同士の相関係数の平均値が 0.63 と相関が見られたが、Blog では 0.37 と弱い相関となった。これはヒアリング調査の結果、ニュース映像では、“アナウンサのまとめの部分をニュースの主旨として評価した”という意見が多かったのに対し、Blog では、“エントリを読んで評価した”、“コメントの部分から判断した”など、Blog の評価割合は各個人の読み方に依存し、注目している部分が異なるためであると考えられる。

ヒアリング調査では、対応付けを行う際の基準は、“話題が一致しており、評価としても近いもの”という基準であったが、ニュース映像と対応付けられた Blog に関して、各被験者で共通する項目はほとんど見られなかった。これは、ニュース映像で語られているどの部分に対して対応付けを行うかという基準が、各個人で異なったためであると考えられる。たとえば「沖縄米軍基地の問題」であれば、被験者によって知事に関する評価を述べている Blog を対応付ける、もしくは米軍基地自体に対する評価を述べている Blog を対応付けるというように、被験者ごとにニュース映像から読み取った話題の中心が異なるためと考えられる。

#### 4.2.4 実験 4：Blog 検索の精度

提案手法の対象語と評価割合を用いた検索の評価を行うため、実際に Blog 検索を行った際の適合率、相対的再現率 (relative recall) <sup>13)</sup>、F 値を求めた。相対的再現率は Henzinger ら<sup>13)</sup> によって、“生成されたすべての検索質問の結果から得られた正解より集合を生成し、それを全体の正解集合とし再現率を算出する手法”と定義されている。実験に用いたニュースは、表 1 のうち、ニュース番号 1 から 5 の 5 件を用い、検索結果として用いた Blog は、各質問に対し gooBlog を用いて検索し、得られた Blog の重要度計算結果の上位 50 件を用いた。以下に実験の手順を示す。

- (1) 被験者は、視聴したニュース映像から抽出され

対応付けられた Blog が複数個の場合はその評価割合の平均値を使用した。  
相対的再現率を求めるために、提案手法以外に、頻出単語で検索した結果も取得し、被験者に正解を判定させた。

表 3 Blog 検索の精度

Table 3 Experimental result of Blog search.

ニュース番号	Precision	Relative recall	F-measure
1	0.389	0.636	0.483
2	0.238	0.370	0.290
3	0.490	0.714	0.581
4	0.571	0.632	0.600
5	0.500	0.563	0.529

た対象語により検索された Blog を閲覧する．

- 生成された評価割合のうち、最も肯定のもの、最も否定のもの、および中立のもの 3 種の検索結果を用いる．
  - 提示する Blog はニュースから得られた評価割合近傍の評価割合となる 3 件を提示する．
- (2) 閲覧した Blog それぞれに対し評価を行う．
- ニューストピックと適合しかつ、生成された評価割合とも合致する場合は 1 点．
  - ニューストピックと適合しているのみの場合、あるいは無関係の場合は 0 点．
- (3) 得られた Blog の評価点数の合計を評価点数の最高点の合計で除算した値を適合率として算出する．
- (4) 得られた Blog の評価点数の合計を提案手法の正解と頻出単語による正解の件数で除算した値を相対的再現率として算出する．

実験結果を表 3 に示す．表中の Precision は適合率であり、Relative recall は相対的再現率、F-measure は F 値である．以下に考察を述べる．

- インタビューや記者会見のような素材をメインにニュースが構成されている場合に、検索精度が低下する結果となった．表 1 のニュース番号 2 では、総裁候補の 3 人が演説するシーンを中心に構成されている．このような場合、対象となる演説を行っている人物を表すようなキーワードは音声テキストに現れにくく、システムは演説者が述べている内容から対象語を抽出するため、検索される Blog もニュースの対象と異なる Blog が検索されてしまう．このような場合、映像中に出現するテロップを手がかりに対象を特定する手法で補足できるものと考えられる．
- 全体として、相対的再現率のほうが高い結果となった．対象語による様々な評価を得ることを目的とするため、可能な限りトピックの範囲を広げるキーワード抽出を行うべきであると考えられる．相対的再現率が高い結果となっているのは、正解

の幅を広げるようなキーワードで検索を行っているためと考えられる．たとえば、表 1 のニュース番号 5 であれば、「村井、長野、田中」というキーワードが抽出されており、前知事の田中氏と新知事の村井氏についての評価が述べられている Blog が検索されている．

#### 4.2.5 実験 5：従来の評判情報検索との比較

提案手法の特性を考察するために従来の評判情報検索である blogWatcher<sup>26)</sup> との比較を行った．実験の手法として、提案手法で得られた Blog と blogWatcher の評判検索に対して提案手法の対象語を入力した際に得られた出力の比較を行った．例として用いたニュースは FNN の 2006 年 5 月 25 日「靖国神社への合祀の取り消しなどを求めている裁判で韓国人戦没者遺族ら全面敗訴」および TBS の 2006 年 5 月 24 日「麻生外相、アジア外交に自信のぞかせる」というニュースである．提案手法、blogWatcher とともに対象語として得られた「靖国神社」、「麻生」というキーワードを用いて検索を行った．その結果を表 4 に示し、得られた類似点および相違点を以下に述べる．

両手法の類似点は、入力されたキーワードについて Blog を肯定・否定に基づいての分類を行い提示する点である．相違点は、評価の解析において、blogWatcher では入力キーワードがどのようなキーワードで評価をされているかを解析しているのに対し、本手法ではエントリのみではなく、コメントも含めた議論の流れでどう評価されているのかを考慮している点である．たとえば、blogWatcher では「麻生」に対して「好きだ」というキーワードでポジティブに評価されているが、本手法では、「麻生大臣は好き」と書かれている Blog でも「外務省を変えられていない」という批判の中で述べられているため Blog で述べられている評価としては否定的であると判定する．また、提示方法では、blogWatcher では評価を行っている Blog 件数の時間経過にともなう推移を提示することが可能であり、本手法では評価割合をユーザが選択することで、1 つの Blog でどのような度合いで評価されているかを閲覧することが可能である．blogWatcher による評判情報検索は、具体的な商品名や人物名の評判を知りたい際には有効であるといえるが、トピックのように様々な見方のできるものに対しては不向きであると考えられる．これらのことより特に、以下の場合に本手法が有効であると考えられる．

blogWatcher では入力された 1 語のみに対する評判情報を得ることを想定している．条件を同一にするために用いる対象語を 1 語とした．

表 4 従来の評判情報検索との比較  
Table 4 Experimental result of opinion information retrieval.

入力 キーワード	blogWatcher		提案手法
	ポジティブな評価	ネガティブな評価	Blog の内容
靖国神社	空気 興味 心地 いい 自由だ 尊い	時代 手間 コース イヤなもの 危険だ 惜しい	・靖国神社が中国語訳や韓国語訳のパンフレットを準備して相互理解に動き出している ・靖国参拝自体が外交問題であり、選挙の争点となるかどうかという問題ではない
麻生	顔 動揺っぷり 時間 イイ 好きだ かわいい	意味 目 無益であると しんどい 悲しい	・外務省に対する鈴木宗雄の質問への麻生大臣の答弁への批判 ・中国への援助に対する麻生大臣の位置づけ

- ・ ニュースより自動的にキーワードを抽出し、ニューストピックに対する評価を得たい場合
- ・ 肯定、否定の度合いを指定した Blog の閲覧を行いたい場合

blogWatcher はキーワードに関する直接的な評価を得たい場合に有効であり、本手法とは利用目的が異なるため、補完的な関係であるといえる。

## 5. おわりに

本稿では、ニュース映像の評価と Blog の評価の統合方式を提案した。対象語と評価割合からなる評価の構造を定義し、プロトタイプシステムの作成および評価実験を行った。提案手法は大きく 2 つの部分からなる。1 つはニュース映像のシーン順序より評価を抽出する部分、もう 1 つは評価に基づく Blog を検索する部分である。ニュース映像のシーン順序より評価を抽出する部分では、映像のシーン内での音声テキストの特徴と編集されたシーン順序の特徴を利用したキーワードの重要度付けを行った。また、評価の度合いに関しては、キーワードの重要度を用いて評価割合という数値を算出することで、Blog で使用されるキーワードとの差異を緩和した。評価に基づく Blog の検索部分では、Blog 中に出現するキーワードの役割による重要度を考慮した検索を行い、評価割合別にニュース映像との対応付けをすることで、ニュース映像で述べられている評価についての Blog を検索することを実現した。実験では以下のことを示した。

- ・ 映画におけるクレショフ効果のようにニュース映像のシーン順序によりユーザの印象の評価が変化すること
- ・ シーン順序に基づくニュース映像の評価割合算出、個数を用いた Blog の評価割合算出が行えること
- ・ 提案手法によりニュースのトピックの内容を含み、評価を行っている Blog を検索できること

今後の課題としては、音声テキストに含まれないが対象として重要なキーワードの抽出、Blog の評価割合算出方式の検討、ユーザに評価を提示する際の感性

語の選択手法とその実装、大規模な実データに基づく実験があげられる。

謝辞 本研究の一部は、平成 18 年度科研費基盤研究 (B) (2) 「Web アーカイブと映像アーカイブを融合した次世代デジタル・ライブラリに関する研究」(課題番号: 16300028) によるものです。ここに記して謝意を表します。

## 参考文献

- 1) CNN.com International. <http://edition.cnn.com/>
- 2) CBSNews.COM. <http://www.cbsnews.com/>
- 3) ABC News. <http://abcnews.go.com/>
- 4) ANN NEWS. <http://www.tv-asahi.co.jp/ann/news/web/>
- 5) FNN-NEWS.COM. <http://www.fnn-new.com/>
- 6) Chen, H.-W., Kuo, J.-H., Chu, W.-T. and Wu, J.-L.: Action Movies Segmentation and Summarization Based on Tempo Analysis, *Proc. 6th ACM SIGMM International Workshop on Multimedia Information Retrieval (ACM MM MIR '04)*, pp.251–258 (2004).
- 7) Kitayama, D. and Sumiya, K.: An Evaluation System for News Video Streams and Blogs, *Proc. 21st ACM Symposium on Applied Computing (SAC 2006)*, pp.361–368 (2006).
- 8) 中島義明: 映像の心理学, サイエンス社 (1996).
- 9) Braudy, L. and Cohen, M.: *Film Theory and Criticism: Introductory Readings*, Oxford University Press (1999).
- 10) TBS NEWS i. <http://news.tbs.co.jp/>
- 11) 日テレ News24. <http://www.news24.jp/>
- 12) Ma, Q., Sumiya, K. and Tanaka, K.: WebTelop: A Dynamic TV-content Augmentation by Using Web Pages, *Proc. IEEE International Conference on Multimedia & Expo (ICME2003)*, Vol.2, pp.173–176 (2003).
- 13) Henzinger, M., Chang, B.-W., Milch, B. and Brin, S.: Query-Free News Search, *Proc. 12th International World Wide Web Conference (WWW2003)*, pp.1–10 (2003).

- 14) Nanno, T., Suzuki, Y., Fujiki, T. and Okumura, M.: Automatic Collection and Monitoring of Japanese Weblogs, *Proc. WWW2004 Workshop on the Weblogging Ecosystem: Aggregation, Analysis and Dynamics*, pp.320–321 (2004).
- 15) Fukuhara, T., Murayama, T. and Nishida, T.: Analyzing concerns of people using Weblog articles and real world temporal data, *Proc. WWW 2005 2nd Annual Workshop on the Weblogging Ecosystem: Aggregation, Analysis and Dynamics* (2005).
- 16) Glance, N.S., Hurst, M. and Tomokiyo, T.: BlogPulse: Automated Trend Discovery for Weblogs, *Proc. WWW 2004 Workshop on the Weblogging Ecosystem: Aggregation, Analysis and Dynamics* (2004).
- 17) Kumuar, R., Novak, J., Raghavan, P. and Tomkins, A.: On the Bursty Evolution of Blogspace, *Proc. 12th International World Wide Web Conference (WWW2003)*, pp.568–576 (2003).
- 18) Gruhl, D., Guha, R., Liben-Nowell, D. and Tomkins, A.: Information Diffusion Through Blog-space, *Proc. 13th International World Wide Web Conference (WWW2004)*, pp.491–501 (2004).
- 19) Dave, K., Lawrence, S. and Pennock, D.M.: Mining the peanut gallery: Opinion Extraction and Semantic Classification of Product Reviews, *Proc. 12th International World Wide Web Conference (WWW2003)*, pp.519–528 (2003).
- 20) 立石健二, 石黒義英, 福島俊一: インターネットからの評判情報検索, 情報処理学会研究報告, Vol.2001-NL-144-11, pp.75–82 (2001).
- 21) Chaisorn, L., Chua, T.-S. and Lee, C.-H.: Extracting Story Units in News Video, *Proc. International Workshop on Advanced Image Technology 2003 (IWAIT 2003)* (2003).
- 22) Ide, I., Yamamoto, K., Hamada, R. and Tanaka, H.: An automatic video indexing method based on shot classification, *Systems and Computers in Japan*, Vol.32, No.9, pp.32–41 (2001).
- 23) Asahara, M. and Matsumoto, Y.: Extended Models and Tools for High-performance Part-of-Speech Tagger, *Proc. 18th International Conference on Computational Linguistics (COLING 2000)*, pp.21–27 (2000).
- 24) Yahoo!Blog. <http://blogs.yahoo.co.jp/>
- 25) gooBlog. <http://blog.goo.ne.jp/>
- 26) blogWatcher. <http://blogwatcher.pi.titech.ac.jp/>

(平成 18 年 9 月 20 日受付)

(平成 19 年 1 月 10 日採録)

(担当編集委員 中挾 知延子)



北山 大輔 (学生会員)

2005 年姫路工業大学環境人間学部環境人間学科卒業。現在、兵庫県立大学大学院環境人間学研究所博士前期課程に在籍。Web 検索、映像データベースに興味を持つ。日本データベース学会学生会員。



角谷 和俊 (正会員)

1988 年神戸大学大学院工学研究科修士課程修了。同年松下電器産業(株)入社。ソフトウェア開発環境、マルチメディアデータベース、データ放送の研究開発に従事。1998 年神戸大学大学院自然科学研究科博士後期課程(情報メディア科学専攻)修了。1999 年神戸大学都市安全研究センター都市情報システム研究分野講師, 2000 年同助教授。2001 年京都大学大学院情報学研究所社会情報学専攻助教授。2004 年兵庫県立大学環境人間学部環境人間学科教授, 現在に至る。博士(工学)。IEEE Computer Society, ACM, 電子情報通信学会, 日本データベース学会等各会員。