

# ニュース構成パターンに基づくビデオ・アーカイブコンテンツ閲覧方式

北山 大輔<sup>†</sup> 角谷 和俊<sup>††</sup>

<sup>†</sup> 兵庫県立大学大学院環境人間学研究科 〒 670-0092 兵庫県姫路市新在家本町 1 丁目 1-12

<sup>††</sup> 兵庫県立大学環境人間学部 〒 670-0092 兵庫県姫路市新在家本町 1 丁目 1-12

E-mail: <sup>†</sup>hd05r011@stshse.u-hyogo.ac.jp, <sup>††</sup>sumiya@shse.u-hyogo.ac.jp

あらまし 現在, TV やインターネットなどを通して映像ニュースコンテンツが配信されている. 一般にニュースは時間が経過すると価値が無くなると考えられるが, 関係するニュースと同時に閲覧することで過去のコンテンツであっても価値が生じる場合がある. 例えば, オリンピックなど何度も起こる類似のイベントにおける前回のメダル獲得時のニュースなどの場合である. そこで本稿では, ニュース映像の構成要素からなるニュース構成パターンを定義し, 構成パターンより関係を特定する手法を提案する. また, 得られた関係に応じてニュースコンテンツの位置関係や再生時間を変化させ提示を行う手法を提案する.

キーワード 情報統合, Web とインターネット, ニュースアーカイブ, ニュース映像

## Video Archive Contents Browsing Method based on News Structure Patterns

Daisuke KITAYAMA<sup>†</sup> and Kazutoshi SUMIYA<sup>††</sup>

<sup>†</sup> Graduate School of Human Science and Environment, University of Hyogo

1-1-12 Shinzaike-honcho, Himeji, Hyogo 670-0092, Japan

<sup>††</sup> School of Human Science and Environment, University of Hyogo

1-1-12 Shinzaike-honcho, Himeji, Hyogo 670-0092, Japan

E-mail: <sup>†</sup>hd05r011@stshse.u-hyogo.ac.jp, <sup>††</sup>sumiya@shse.u-hyogo.ac.jp

**Abstract** Recently, video news contents are broadcast on both TV and the Internet, Out of date news video contents have no value of viewing, however, the contents are considered of value by comparing the related contents. Especially, the repeated news should be compared, for example Olympic games, international expositions and so on. We propose a retrieval method of related contents based on news structure patterns. Furthermore, we develop a browsing method which determine the showing position of the screen and playing schedule according to the calculated relation.

**Key words** Information integration, Web and Internet, News Archive, News Video Stream

### 1. はじめに

映像メディアによる情報取得は閲覧の容易性から TV のみならずインターネットにおいても一般的となってきた. 特にニュースは情報源として重要であり, インターネット上でも FNN-NEWS.COM<sup>(注1)</sup>, TBS News<sup>(注2)</sup>, ANN NEWS<sup>(注3)</sup>, 日テレ NEWS24<sup>(注4)</sup> といった各報道局により映像ニュースサイトが公開されている. しかし, 映像ニュースサイトでは映像が公開されている期間は短い. これは, 一般にニュース映像は速報性を重視しているためではないかと考えられる.

一方, 特集番組などの場合, 過去のオリンピック競技の映像と現在の競技の映像を比較しながら映像を構成するというとも行われている. そのため, ニュースとしての価値が失われた映像であっても, 現在の事柄との関係性を示すことで閲覧する価値が生じるのではないかと考えられる. そこで本稿では, 過去のニュース映像コンテンツを閲覧し比較ができると便利であると考え, インターネット上で公開されているニュースコンテンツを収集しアーカイブするサイトにおいて, アーカイブ中のコンテンツを映像として有用に閲覧する手法を提案する. ユーザは閲覧したいニュースを選択することで, 時間的に近いニュースはもちろん, 時間的に遠いニュースであっても選択したニュースとの関係性に基づき閲覧することが可能となる. 以下, 2 節において研究の概要と関連研究について述べ, 3 節ではニュースの関係性の抽出方法について説明し, 4 節で関係性に基

(注1): <http://www.fnn-news.com/>

(注2): <http://news.tbs.co.jp/>

(注3): <http://www.tv-asahi.co.jp/ann/news/web/>

(注4): <http://www.news24.jp/>

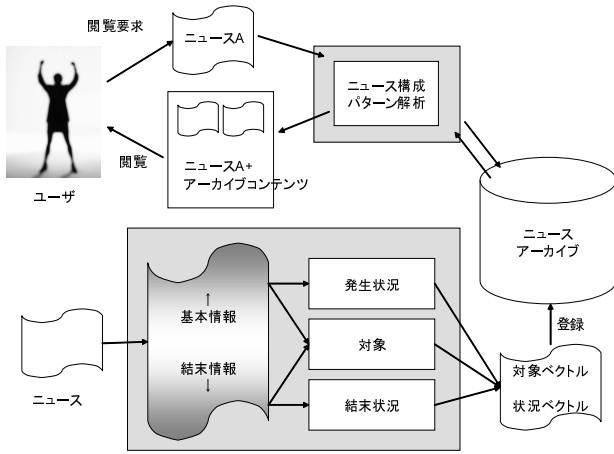


図1 システム概要図

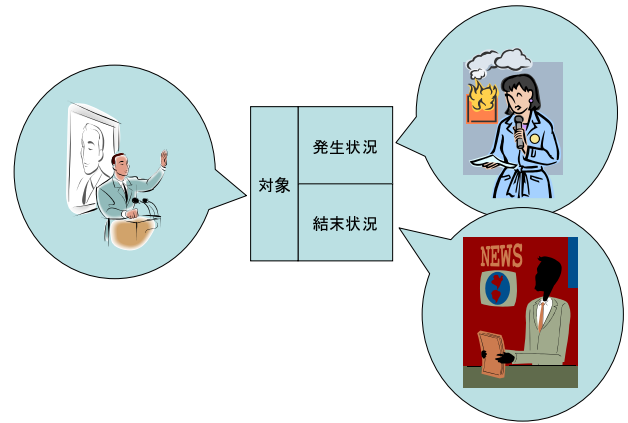


図2 ニュース映像の構成パターン

づく提示方法を述べる。5節でプロトタイプについて述べ、最後に6節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 研究の概要

本稿で提案する閲覧方式では、ニュース映像中のキーワードの出現位置による役割と、役割を考慮したキーワードベクトルからなる構成パターンの類似性によりニュースの関係性を分類し、ニュースコンテンツを提示する(図1)。

ニュース映像では、冒頭と終端でニュースとして基本的な情報を伝え、ニュースの流れとして発生状況、その詳細、状況の結末へと時系列に構成される。単語の繰り返しを用いながら進展するため、この意味単位はシーンという単位とは異なり、明確に区分することができない。これらニュース映像の特性によりニュース映像の構成要素として、ニュースの対象・発生状況・結末状況を定義する。これらは、単語の出現位置により重みが決まるベクトルである。これらの組み合わせを構成パターンと定義する(図2)。また、構成パターンの類似性により、決定される関係として、対比・類比・続報・代替を定義する。これらの関係により、あるニュースコンテンツと他のニュースコンテンツを閲覧する際の閲覧方法を決定し提示を行う。

### 2.2 関連研究

現在提供されているニュースサイトとしてMcKeownら[3]のNewsblaster<sup>(注5)</sup>やRadevら[4]のNewsInEssence<sup>(注6)</sup>やGoogle-News<sup>(注7)</sup>があげられる。これらのニュースサイトは主として、そのトピックを簡潔に理解するために続報記事の集約・要約を目的としており、本稿で提案するアーカイブコンテンツとの関係の提示を行う手法とは目的が異なる。

ニュース映像のトピックを追跡する研究として井手ら[6]の研究があげられる。これは、ニュース映像の字幕テキストを用い、その類似度によってトピック追跡を行う手法である。また、トピックの結合状況に応じて段階的に追跡できるようにスレッ

ド構造を定義している。本手法は、トピックとして連続していないようなニュースとの関係性をも導き出す事を目的としている点でトピックの追跡とは異なる。

ニュースの構成要素の抽出に関する研究として井手ら[5]や戸田ら[7]の研究があげられる。井手らの手法はニュース映像のオープンキャプションやクローズドキャプションよりニュースの構成要素として4W(Who, Where, When, What)に相当するキーワードの抽出を行うものである。戸田らの手法はニュースの構成要素として固有表現に着目し、トピックごとに固有表現の種類に対する重みを変化させることによりクラスタリングの精度を向上させる手法である。また、映像の構成要素を抽出して提示を行うシステムとしてDimitrovaら[2]のInfoSipがあげられる。これは、TV番組の5W1Hといった内容に関する付加情報を提示できるシステムである。本手法において着目しているのは5W1Hや固有表現といった特定のキーワードをニュースの構成要素として捕らえるのではなく、映像中の出現位置による役割に基づいてニュースの構成要素を決定するもので、着眼点が異なる。

ニュースの視覚的な提示手法として、Christelら[1]の研究があげられる。これはニュースビデオを地理的や時間的に並べて配置することで映像の閲覧効率を上げたり、視覚的にニュースの要約を行う手法である。本手法とはニュースの関係によって映像の提示を変えるという点で類似している。しかし、本手法におけるニュースコンテンツの提示は時間や地理的な尺度ではなく、ニュース映像で述べられている内容に基づくものであり、観点が異なる。

閲覧するニュースに加え過去のコンテンツを同時に提供する研究として灘本ら[8]のT-CNBがあげられる。これは閲覧するニュースのトピックを時間経過にそって追っていけるように並列表示を行うものである。本稿で提案する手法は、単純に時間に沿って追うのみではなく、映像の構成の類似性によって表示方式を変えるものである。

アーカイブコンテンツの利用についての研究として吉岡ら[9]の研究があげられる。これは、閲覧しようとするニュースに対してトピックの側面という観点から整理を行い、閲覧ニュースを含むニューストピックがどのように構成されているのかを関

(注5): <http://www1.cs.columbia.edu/nlp/newsblaster/>

(注6): <http://www.newsinsence.com/nie.cgi>

(注7): <http://news.google.co.jp/>

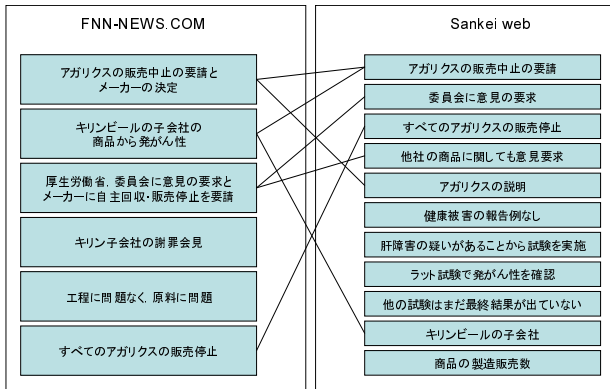


図3 予備実験：アガリクスから発がん性販売中止決定

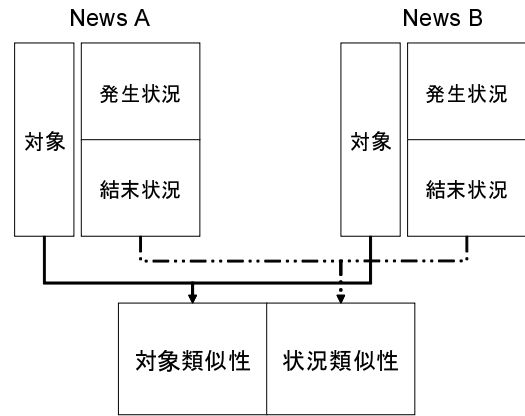


図4 構成パターンの類似性

覧する手法である．本手法は閲覧ニュースそのものに対する他のニュースとの関係性を導き出し閲覧することを目的とするものでありトピックの構成の閲覧とは異なる．

### 2.3 予備実験

同一のニュースを扱っていたとしても、メディアによってその表現の手段が異なると考えられる．テキストのニュースであれば、時系列に関係なく読み返すことができるため、論理的に説明するように構成されていると考えられる．また、映像のニュースは時系列的に理解するため、繰り返しを用いながら一つのストーリーを展開すると考えられる．映像メディアによるニュースの伝達の特徴を明確にするための予備実験を行った．手順は以下のとおりである．

1. 映像ニュースに対応するテキストのニュースを選択する．
2. 映像ニュースの音声テキストより、文を時系列順に並べる．
3. テキストニュースより、文を文書順に並べる．
4. 対応するニュースの文の出現順の対応付けを行う．

実験に用いた映像ニュースはFNN-NEWS.COM, ANN NEWS, テキストニュースはSankei Web<sup>(注8)</sup>, asahi.com<sup>(注9)</sup>でそれぞれ2月14日のものを用いた．これは、FNN-NEWS.COMはSankei Web, ANN NEWSはasahi.comに相当すると考えられるからである．これらのニュースサイトより対応する7件の記事を選択し、それぞれの文の対応づけを行った．例として取り上げた一つを図3に示す．結果は以下のとおりである．

- 映像中では前半の内容がテキスト中では後半：2件
- 映像中では後半の内容がテキスト中では前半：7件
- 映像中のみにはしかない内容が存在：4件

図3の左側の四角は一つの映像ニュースの音声テキストを表し、右側の四角は一つのテキストニュースを表す．大きな四角中の長方形はそれぞれに含まれる文の要約であり、それらをつなぐ線は対応する文であることを示す．線のないものは、対応する文がないと判断したものである．

この結果より、ニュースで述べられる内容の出現傾向として、映像中前半に出現する内容はテキストでも前半で扱われており、また映像では後半に位置するが、テキストでは前半に出現するような文が存在するという傾向が見られた．この映像では後半、テキストでは前半に述べられる文は、図3の例では“すべての商品の販売を中止する”といったようにニュースの発生後の結末に相当する内容であると考えられる．映像にしか出現しない内容、テキストにしか出現しない内容といったものも見られ、映像では中心付近に、テキストでは後半に出現するという傾向が見られた．

一般にテキストのニュースでは、はじめに概要を述べ、次に詳細、続いて補足的な内容について述べられるといわれる．このことより、テキストにしか出現していない内容は補足的な内容であると考えられる．また映像にしか出現しない内容は、結末への流れを作るための内容であると考えられる．図3の例では“記者会見での謝罪”を入れ、前半と後半に一区切り入れることで後半の“原料自体が問題”、“すべての商品の販売を中止する”を導きやすくしていると考えられる．

この予備実験により2つの仮説を得た．一つは映像ニュースではニュースを表す基本的な内容は冒頭と終端に分散し、その内容はニュースの発生とその結末であること．もう一つは、ニュース映像では次のシーンを導くために前のシーンを構成しているということである．

## 3. 構成パターンによるニュース間の関係の抽出

### 3.1 ニュース間の関係

本稿ではニュース間の関係を2つのニュース構成パターンの類似・非類似のパターンによって分類を行う(図4)．ニュース構成パターンとは、ニュース対象・発生状況・結末状況の3種類の構成要素の組み合わせである．また、これらによって得られるニュース間の関係として対比関係、類比関係、続報関係・代替関係の4つを定義する<sup>(注10)</sup>．以下にニュース間の関係とそれを決定する構成要素について述べる．

- 対比関係

(注8) : <http://www.sankei.co.jp/>

(注9) : <http://www.asahi.com/>

(注10) : これらの関係に時間的な周期性、一時性、前後、ランダムといったものを加味すること考えられるが、本稿では取り扱わない．

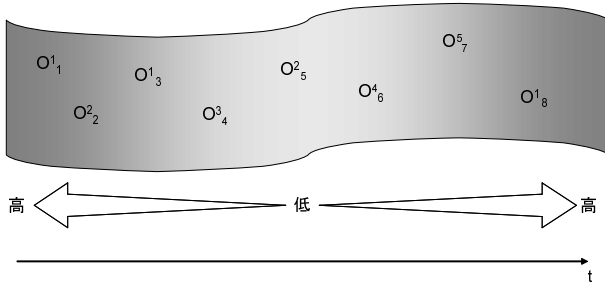


図5 対象重要度の算出

2つのニュースの異なる点に着目して閲覧することが有用であると考えられる関係である。例えば、万博の開幕と万博の開幕など、対になるような場合である。このような関係になるニュースは、対象が類似し、発生状況や結末状況が相互に類似すると考えられる。

- 類比関係

2つのニュースの類似点に着目して閲覧することが有用であると考えられる関係である。例えば、楽天・TBSの統合問題と村上ファンド・阪神電鉄の買収問題といった、2つの出来事に直接の関係はないが比較したいような場合である。このような関係になるニュースは、対象が非類似であり、発生状況や結末状況がそれぞれ類似すると考えられる。

- 続報関係

2つのニュースが続きの報道である場合の関係である。続報として閲覧する場合、話の進展があるものを閲覧する方が良いと考えられる。つまり、同じ対象についてのニュースであるが異なる事態がわかったために続きとしてニュースが発せられたものを続報関係とする。そのため続報関係になるニュースは、対象が類似するが、発生状況・結末状況が類似しないと考えられる。

- 代替関係

2つのニュースが同じ対象、同じ状況にあり相互に入れ替えても問題のない関係である。例えば、異なる局の同じニュースであったり、毎年行っているイベントというような場合である。このような関係になるニュースは、対象や発生状況・結末状況が類似すると考えられる。

### 3.2 構成要素の抽出

本節ではニュース映像の構成要素の抽出方法について説明する。ニュースの対象・発生状況・結末状況は単語ベクトルとして抽出を行う。ニュース中で対象は一般名詞・固有名詞で表現されていると定義し、また、発生状況・結末状況はサ変名詞・自立動詞で表現されていると定義する。

まず、ニュースの対象の抽出について述べる。ニュース映像において、ニュースを構成する対象は冒頭と終端部分に出現しやすいと考えられる。このため、冒頭、終端に近いほど重要とするような関数を用いて対象重要度  $val_o$  を求める(図5)。この対象重要度をニュースの名詞からなる特徴ベクトルの各要素ごとに乗算することで対象ベクトル  $O$  を生成する。

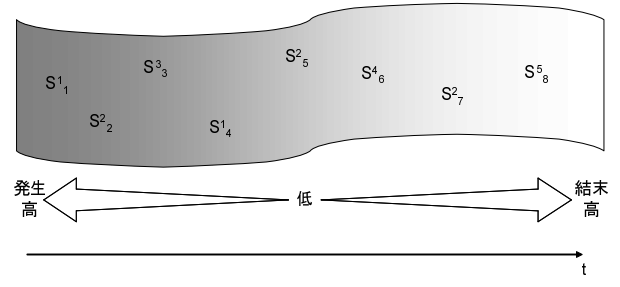


図6 状況重要度の算出

$$val_{o^{id}} = \sum_{i=0}^n \begin{cases} \exp(-\frac{n}{(i-n/2)^2+1}) & (o^{id} = o_i) \\ 0 & (o^{id} \neq o_i) \end{cases} \quad (1)$$

$$O = v(N_{nom}) \times val_{o^{id}} \quad (2)$$

式中の  $o^{id}$  はニュース中に出現する名詞であり、 $id$  は特徴ベクトルの該当箇所に対応する。名詞は形態素解析により、一般名詞、固有名詞を抽出する。 $n$  はニュース中に出現する名詞数、 $i$  は出現順である。 $v(N_{nom})$  はニュース中の名詞からなる特徴ベクトルである。

次に、ニュースの発生状況、結末状況の抽出について述べる。ニュース映像において、ニュースの起こりを示す発生状況は冒頭部分に、ニュースの今後を示す結末状況は終端部分に出現しやすいと考えられる。このため冒頭に近いほど発生状況として重要、終端に近いほど結末状況として重要となるような関数を用いて状況重要度  $val_s$  を求める(図6)。この状況重要度をニュース中の動詞からなる特徴ベクトルの各要素ごとに乗算することで状況ベクトル  $S$  を生成する。

$$val_{s^{id}} = \sum_{i=0}^n \begin{cases} \exp(-\frac{n}{(i-n/2)^2+1}) & (o^{id} = o_i \wedge i < n/2) \\ -\exp(-\frac{n}{(i-n/2)^2+1}) & (o^{id} = o_i \wedge i \geq n/2) \\ 0 & (o^{id} \neq o_i) \end{cases} \quad (3)$$

$$S = v(N_{verb}) \times val_{s^{id}} \quad (4)$$

式中の  $s^{id}$  はニュース中に出現する動詞であり、 $id$  は特徴ベクトルの該当箇所に対応する。動詞は形態素解析により、動詞、サ変名詞を抽出する。 $n$  はニュース中に出現する動詞数、 $i$  は出現順である。 $v(N_{verb})$  はニュース中の動詞からなる特徴ベクトルである。 $val_{s^{id}}$  はニュースの発生状況にかかわる場合は正の値、ニュースの結末状況にかかわる場合は負の値をとる。

抽出されるベクトルとして、“野菜小売価格、雪害の影響でキャベツは例年に比べ54%高 農水省調査”(図7)のニュースの場合であれば、対象ベクトルの重要度上位の単語として{野菜, 雪害}といったキーワード、状況ベクトルの発生状況として{小売, 比べる}, 結末状況として{目立つ, 高騰}といったキーワードの重要度が高くなる。これは、例年と比較して野菜の値段が高くなったことによって発生し、葉物野菜の高騰が目立つので調査を行うという結末であるニュースとなり、ニュース本文から正しく構成要素を抽出できているといえる。

大雪で被害が出ている野菜などの小売価格の調査を農林水産省が行った結果、キャベツは例年に比べて5割も高くなっていることがわかった。農水省によると、1月第1週の野菜の小売価格は、過去5年間の平均値に比べ、キャベツが54%高、ネギが42%高、ホウレン草が40%高など葉物野菜の高騰が目立っているという。農水省は、雪害による価格動向を把握するため、今後も毎週調査を実施するという。

図7 各重要度の高くなる単語の例

### 3.3 構成パターンの比較とニュースの関係の決定

2つのニュース間で構成パターンの比較について述べる。対象と状況に関してはベクトル空間によるコサイン相関値により比較を行う。関係性を表す指標として対比度 *Comparison*、類似度 *Analogy*、続報度 *Follow-up*、代替度 *Substitute* を定義する。これらは、構成パターンの関係性を求める式によって定義される。それぞれの値が閾値を超えている場合に該当する関係であると判断する。そのため、あるニュースはすべての関係に対しニュースが存在するが、別のニュースは関係がまったく存在しないこともある。

$$Comparison = Rel(O_A, O_B) \times -Rel(S_A, S_B) \quad (5)$$

$$Analogy = (1 - Rel(O_A, O_B)) \times Rel(S_A, S_B) \quad (6)$$

$$Follow-up = Rel(O_A, O_B) \times (1 - |Rel(S_A, S_B)|) \quad (7)$$

$$Substitute = Rel(O_A, O_B) \times Rel(S_A, S_B) \quad (8)$$

$$Rel(A, B) = \frac{A \cdot B}{|A| \cdot |B|} \quad (9)$$

式中の  $A, B$  はそれぞれ比較するニュースコンテンツであり、 $O_A$  は  $A$  における  $O$  を表す。 $Rel$  関数により関係性の算出を行う。 $Rel$  関数の値は 1.0 から -1.0 の値をとり、負の値をとるような関係は、ニュース映像の構成が逆になっていることを示す。これらの値はその関係性を表しており、値が大きいほど該当する関係にあるといえる。例えば、ニュース  $A$  に対して  $Comp$  が閾値を超えるニュース  $B$  は対比関係にあると判断する。

## 4. 関係に基づくコンテンツ提示

### 4.1 コンテンツ提示方法

本手法では、ユーザが閲覧を望むコンテンツに対して、ニュースとして付加して提示する価値があるコンテンツをシステムが選択し、再構成を行い提示する(図8)。再構成方法を決定する要因として、構成パターンの対象関係性・状況関係性を用いる。

まず、対象関係性によって決定する要因について述べる。対象関係性によって、並列再生を行うか、連続再生を行うかが決定される。対象関係性が高い場合、これはニュースの基本となる情報が類似していることを示すので、映像として映っているものも類似すると考えられる。そのため対象関係性に応じて、冒頭・終端を並列再生を行う。並列再生を行う時間  $syn\_time$  は、対象関係性の値に比例し、最大で冒頭・終端から映像再生

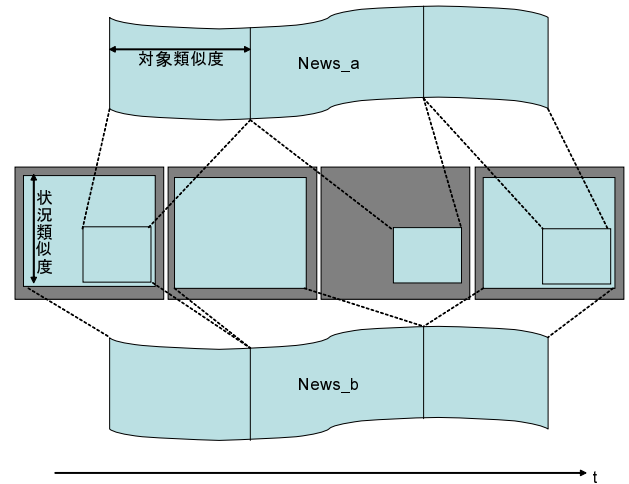


図8 構成パターンに基づくコンテンツ提示

時間  $V\_time$  の  $t$  分の一とする。

$$syn\_time = \alpha \times Rel(O_A, O_B) \times \frac{V\_time}{t} \quad (10)$$

並列再生区間では、ユーザが選択したコンテンツの音声を再生し、システムが選択したコンテンツは字幕再生を行う。連続再生区間では、システムが選択したコンテンツを先に再生し、ユーザが選択したコンテンツを後に再生する。 $\alpha$  は任意の重みである。

次に、状況関係性によって決定する要因について述べる。状況関係性の絶対値によって、再生サイズと再生位置の決定を行う。状況関係性は、ニュースの発生やその結末が類似していることを表す。つまり同じような類のニュースであるということユーザに視覚的に伝えることができれば有効であると考えられる。そのため、状況関係性が高ければ映像がオーバーラップするように配置し再生サイズを大きくする、低ければオーバーラップしないように配置する。再生サイズを大きくするのは、単純にオーバーラップさせると、コンテンツが隠れてしまい閲覧できなくなるためである。再生サイズは、ユーザが選択したコンテンツサイズを  $\beta$  として、システムが選択したコンテンツサイズを変更する。

$$Overlap = \beta - (\beta \times \alpha \times Rel(S_A, S_B)) \quad (11)$$

$$Size = \beta \times \alpha \times (1 + Rel(S_A, S_B)) \quad (12)$$

$Overlap$  はコンテンツの右端の距離を示し、 $Size$  は  $\beta$  に対するサイズ比である。また、状況関係性の値の正負によって、並列再生部分の冒頭・終端を逆転させる。状況関係性が負である場合というのは、逆の状況にあることを示している。そのため、並列再生部分では、比較をしやすくするためにシステムが選択したコンテンツの冒頭と結末を入れ替えて再生する。

### 4.2 ニュース間の関係とコンテンツ提示

#### ● 対比関係

前節で述べたコンテンツ提示方法に基づいて各関係での提示例を述べる。対比として選択されるニュースは、対象が類似し、状況が負で類似する場合である。そのため冒頭・終端が逆になって並列再生され、再生位置がオーバーラップして再生される。



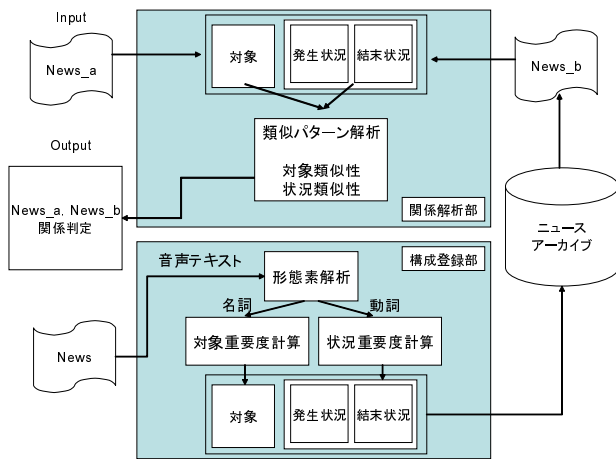


図9 システム構成図

このことにより、ユーザが閲覧するニュースにあわせて、冒頭・中盤・終端が対比しやすいように再生されると考えられる。

- 類比関係

類比として選択されるニュースは、対象が類似せず、状況が正で類似する場合である。そのため並列再生される部分は少なく、映像位置はオーバーラップして再生される。このことにより、似たような類のニュースであることを先に提示し、その後閲覧することができ、同じ種類のニュースであることがわかりやすく再生できると考えられる。

- 続報関係

続報として選択されるニュースは、対象が類似し、状況は類似しない場合である。そのため冒頭・終端は並列再生され、映像位置はオーバーラップしない。このことにより、続報であるための同じ基本情報は省略して閲覧でき、内容として進展した部分を閲覧することができる。

- 代替関係

代替として選択されるニュースは、対象が類似し、状況も正で類似する場合である。そのため冒頭・終端は並列再生され、映像位置はオーバーラップして再生される。このことにより、ほぼ同じニュースであるが、内容の部分を順に再生されることでその小さな差異をじっくり閲覧することが可能であると考えられる。

## 5. 評価

### 5.1 プロトタイプシステム

プロトタイプシステムは大きく分けて構成登録部、関係解析部の2つからなる(図9)。構成登録部では、ニュース映像より構成要素として対象ベクトル、状況ベクトルを生成しニュースアーカイブへの登録を行っている。関係解析部では、ユーザの選択したニュースの構成要素をアーカイブ中のニュースの構成要素と関係性計算を行い、各関係での構成パターンが一致するニュースを検索し、各関係の重要度の高いものをユーザに提示する。このとき構成要素の関係性に基づき、SMILコンテンツを生成する。ユーザはシステムの提示する関係のうち、任意のものを選択することでニュース間の関係に合ったコンテンツ



図11 SMILコンテンツの再生画面：類比関係

を閲覧することができる。

インターフェースとしては、図10のように実装した。左上のブラウザによってユーザはニュースサイトを閲覧することができる。ユーザの閲覧が切り替わるごとに、画面下部のそれぞれの対応する関係をシステムが計算しアーカイブ内のニュースのタイトルが出力する。このリストからユーザは閲覧したいニュースを選択することで右上のブラウザにアーカイブ内のニュースが出力される。中央下部の再生ボタンを押すことでブラウザに出力されている2つのニュースの関係に基づきニュース映像のSMILコンテンツが生成され、プレイヤーで再生される(図11)。画面左下の小さな映像がユーザの選択したニュース映像であり、右上の大きな映像がシステムの選択した映像である。この例は類比関係の場合であり、状況関係性が0.42の場合の画面位置となっている。

プロトタイプで用いたニュース映像はFNN-NEWS.COM、TBS NEWS iの映像ニュースサイトより取得した実際のニュース映像である。構成登録部、関係解析部はVisual Studio.NETのC#により作成した。ニュース映像の音声テキストの単語抽出には茶筌<sup>(注11)</sup>による形態素解析を用い抽出した。

## 5.2 実験

### 5.2.1 実験方法

実験にはFNN-NEWS.COM、TBS NEWS iより取得したニュース映像を用いた。取得期間は2005年8月22日から2006年2月21日までの6ヶ月間(10368件)である。まず、各関係の閾値を求めるための予備実験を行った。予備実験は2005年12月21日から2006年1月20日までの1ヶ月間(2280件)のニュースをサンプルとして用いた。以下に実験の手順を述べる。

1. 無作為に選択したニュースに対し、サンプル内のニュースとの各関係の値を求める。
2. 算出された結果を上位から順に閲覧し、無関係なものが

(注11): <http://chasen.naist.jp/hiki/ChaSen/>

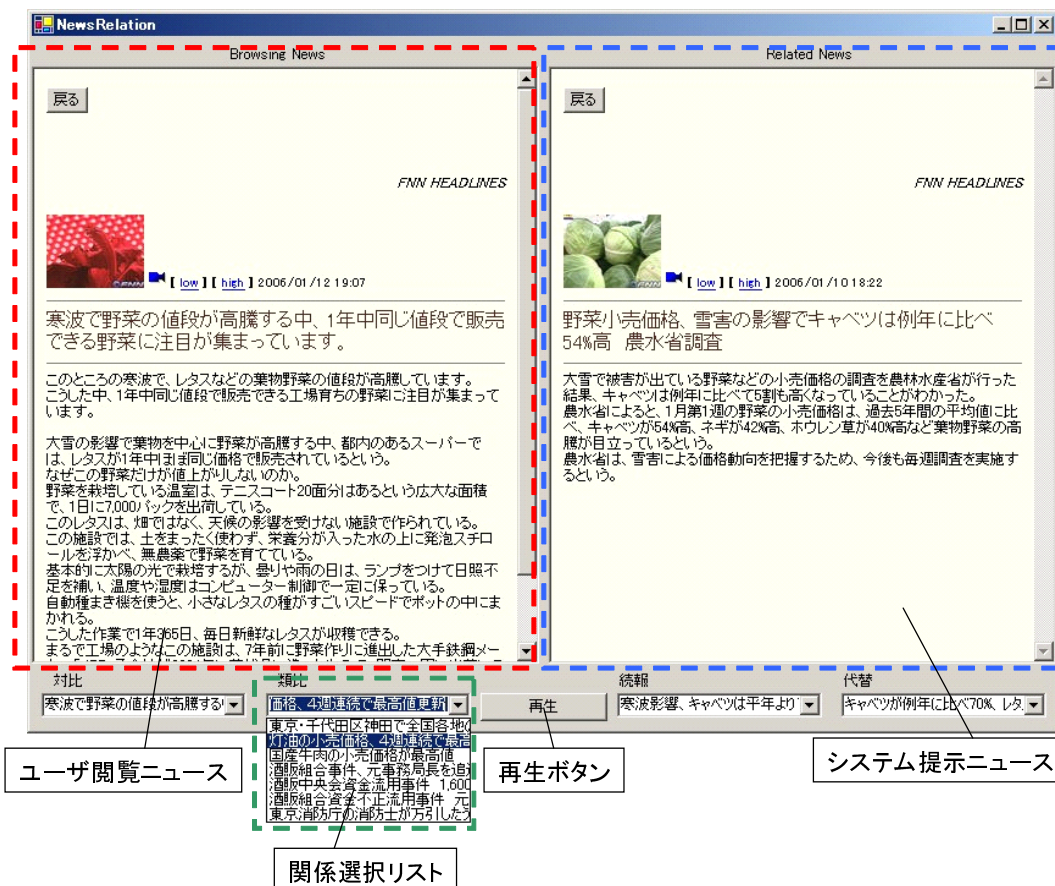


図 10 画面イメージ

出現した関係の値を記録する。

3. 1.2. を各関係に対し 10 件づつ行いその平均値を算出し、関係の閾値とする。

上記の予備実験の結果を以下に示す。実験に用いたニューストピックは、“ライブドア粉飾決算”、“拉致被害者家族会”、“シャロン首相の入院”、“野菜の高騰”といった、社会、政治、国際、経済などさまざまなカテゴリより 10 件選択し実験を行った。これら 10 件によって得られた値のうち、特異なものは出現しなかった。以下に、関係ごとの値の傾向について述べる。

対比関係では 0.023 という値が得られた。対比関係として上位に現れるニュースコンテンツの傾向として、状況関係性の絶対値の低いものが多く見られた。これは、ニュース映像の発生状況と結末状況が綺麗に入れ替わるということが起こりにくく、一部の単語に関して発生したとしても他の単語が正位置で出現するため、絶対値として下がってしまうと考えられる。

類似関係では 0.20 という値が得られた。類似関係として上位に上がるものでは、対象関係性が高いが状況関係性も高いために上位になったコンテンツは類似と呼べる関係ではなかった。

続報関係では 0.24 という値が得られた。続報関係では時間的な特性を考慮していないため、時間的に直接的続報でないものも含まれていた。ただ、これは得られた結果に時間的な重みを付与することで解決すると考えられる。

代替関係では 0.097 という値が得られた。代替関係として、上位に上がるコンテンツでは、同じ局の時間的近傍のあまり

進展のない続報が見受けられた。想定していた他の局の同じニュースというのは表現の違いから、関係性の値があまり高くない傾向が見られた。全体として関係性を掛け合わせているため小さな値となっている。また映像中の出現位置を考慮した重み付けを行っているため、関係性自体の値も低い値となる傾向が見られた。

### 5.2.2 出力結果

得られた閾値を用いて、実際の 6ヶ月間のニュース映像アーカイブに対して各関係ともどのような出力が得られるかという実験を 4 件のニュースに対して行い、適合率の計算を行った。以下に適合であると判断したコンテンツの結果の例示とその考察を示す。表 1 に適合率を示した。また、表 2 は例示したニュースに対して各関係で出力されたニュースの例である。例示するのは“野菜小売価格、雪害の影響でキャベツは例年に比べ 54%高”というニュースに対しての例である。

#### ● 対比関係

対比関係とみなしたコンテンツは、“寒波で野菜の値段が高騰する中、1 年中同じ値段で販売できる野菜に注目が集まっています”といったものである。高騰したことに対して価格を抑える工夫というのは対比関係になっているといえる。出力されているコンテンツの傾向として、対象関係性が低い状況関係性が高いものよりも対象関係性が高く状況関係性の絶対値の低いの方が適合している場合が多かった。これは状況関係性自身が負の値をとりにくいこと、トピックとして関係がない場合、

表1 適合率

	適合率
対比	73.9%
類比	29.6%
続報	57.0%
代替	42.1%

表2 出力されるニュースの例：野菜小売価格、雪害の影響でキャベツは例年に比べ 54%高

	タイトル
対比	寒波で野菜の値段が高騰する中、1 年中同じ値段で販売できる野菜に注目が集まっています
類比	灯油の小売価格、4 週連続で最高値更新
続報	生鮮野菜の小売価格、ほぼ半年並みに
代替	キャベツが例年に比べ 70%、レタスが 55%も値上がり

発生・結末というものの扱いが異なることが考えられる。そのため、同トピックで対比関係にあるものは対象関係性が高く、状況関係性が低めとなると考えられる。

#### ● 類比関係

類比関係とみなしたコンテンツは、“灯油の小売価格、4 週連続で最高値更新”といったものである。野菜の高騰に対して、灯油や牛肉の高騰はトピックとして無関係であるが、よく似たニュースであり類比関係にあるといえる。出力されるコンテンツの傾向として、対象関係性がゼロのものよりも、わずかでも数値のあるニュースの方が適合している場合が多かった。これは、ニュースとして扱っている対象が互いにまったく含まれない場合、何が状況として類似するのかという基準が得られないため関係性を見出すことができないためであると考えられる。

#### ● 続報関係

続報関係とみなしたコンテンツは、“生鮮野菜の小売価格、ほぼ半年並みに”である。野菜の高騰というトピックに対し、状況の異なる場面を出すことができているといえる。そのため、時間的なつながりを考慮しなくとも続報に当たる関係であるといえる。出力されるコンテンツの傾向として、数値的な特徴で適合、不適合という傾向は見ることができなかった。続報に当たるコンテンツというのは、トピックによっては多数存在すると考えられ、閾値外に適合するコンテンツが多数含まれることが予想される。これは閾値の決め方が、はじめに不適合なものが出現した値の平均値であるため適合率重視な閾値となっていることが考えられる。

#### ● 代替関係

代替関係とみなしたコンテンツは、“キャベツが例年に比べ 70%、レタスが 55%も値上がり”である。キャベツの値段の高騰というニュースに対してほぼ同じことを述べているといえる。つまり、このニュースを代わりに提示されてもかまわないニュースであるといえる。出力されるコンテンツの傾向からは、適合・不適合という傾向は見ることができなかった。傾向として、ごく短期間の続報が出力されるか、まったく関係のない良く似たものが出力されるかであった。

## 6. まとめ

本稿ではニュース映像の構成要素を定義し、その要素からなる構成パターンの関係性によって、2 つのニュース間の関係性を分類し、その分類に基づいたコンテンツの閲覧方法の提案を行った。予備実験として映像ニュースとテキストニュースの構成の違い、提案する手法により得られる出力の確認を行った。しかし、評価として十分なデータ量で行うことができていない。また、関係性計算を総当りで行うため、データ量に応じて計算

量が膨大になるという問題点がある。今回、プロトタイプではビューアという形での実装を行ったが、今後は本稿で提案した関係に基づいたニュースアーカイブサイトの構築を行う予定である。今後の課題としては、以下のことを行う予定である。

- 大規模なテキストニュースデータを用いたメディアの差異の比較実験
- 提案手法の精度に関する評価実験
- コンテンツ提示方法に関する評価実験
- 実験の考察によるアルゴリズムの改良
- 大規模データによる実験
- コンテンツ閲覧時のユーザインタフェースの考察
- アーカイブコンテンツ検索方式の提案

## 謝 辞

本研究の一部は、平成 17 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すものとします。

## 文 献

- [1] Christel, M. G., Hauptmann, A. G., Wactlar, H. D. and Ng, T. D.: Collages as Dynamic Summaries for News Video, *Proc. of the tenth ACM international conference on Multimedia*, pp. 561 – 569 (2002).
- [2] Dimitrova, N., Janevski, A., Li, D. and Zimmerman, J.: Who 's That Actor? The InfoSip TV Agent, *Proc. of the 2003 ACM SIGMM workshop on Experiential telepresence*, pp. 76 – 79 (2003).
- [3] Mckeown, K., Barzilay, R., Evans, D., Hatzivassiloglou, V., Klavans, J. L., Nenkova, A., Sable, C., Schiffman, B. and Sigelman, S.: Tracking and summarizing News on a Daily Basis with Columbia's Newsblaster., *Proc. of the Human Language Technology Conference* (2002).
- [4] Radev, D., Otterbacher, J., Winkel, A. and Blair-Goldensohn, S.: NewsInEssence: summarizing online news topics, *Communications of the ACM*, Vol. 48, pp. 95 – 98 (2005).
- [5] 井手一郎, 坂井修一, 田中英彦: ニュース映像中の複数テキスト情報源からの重要語抽出, 第 61 回情報処理学会全国大会講演論文集, Vol. 3, pp. 99–100 (2000).
- [6] 井手一郎, 孟洋, 片山紀生, 佐藤真一: 大規模ニュース映像コーパスの意味構造解析, 電子情報通信学会パターン認識とメディア理解研究会技報, PRMU2003-97 (2003).
- [7] 戸田浩之, 片岡良治, 北川博之: 固有表現を用いたニュース記事分類手法の提案, 情報処理学会研究報告, 2005-DBS-137(I), pp. 175–181 (2005).
- [8] 灘本明代, 田中克己: T-CNB: 時間を考慮した文脈に基づくニュースブラウザの提案, 電子情報通信学会データ工学ワークショップ (DEWS'04) 論文集 (2004).
- [9] 吉岡由智, 湯本高行, 田中克己: ニュースの視点の抽出によるマルチメディアニュースアーカイブの利用, 情報処理学会研究報告, 2005-DBS-137(II), pp. 415–420 (2005).