

プレゼンテーション視聴システムのための異種メディア統合方式

北畑 陽子[†] 平元 綾子^{††} 北山 大輔^{††} 角谷 和俊^{†††}

[†] 姫路工業大学環境人間学部 〒 670-0092 兵庫県姫路市新在家本町 1 丁目 1-12

^{††} 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

^{†††} 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: [†]{na03z056,nd06c019,nd05r011}@stshse.u-hyogo.ac.jp, ^{††}sumiya@shse.u-hyogo.ac.jp

あらまし 現在，ビデオやスライドなどの異種メディアを組み合わせたプレゼンテーションが頻繁に行われるようになってきた．また，蓄積されたスライドやビデオといった各メディアに対し，さまざまな検索手法が提案されている．そこで，本稿では各メディアデータの特性を考慮した検索方式を提案する．本方式では，蓄積されたプレゼンテーションの視聴時において使用するメディアをビデオとスライドとし，ユーザの入力したクエリキーワードがどのメディアに出現するかを考慮することにより，クエリに最も適した区間の検索を行う．これによりクエリキーワードの有無のみに依存するだけでなく，視聴者にとって，より効率的な視聴のできる区間を抽出することができる．

キーワード 情報統合，パーソナライゼーション，情報検索，時空間 DB

An Integration Method using Heterogeneous Media for Multimedia Presentation Systems

Youko KITAHATA[†], Ryoko HIRAMOTO^{††}, Daisuke KITAYAMA^{††}, and Kazutoshi SUMIYA^{†††}

[†] School of Humanities for Environmental Policy and Tecnology, Himeji Institute of Technology
1-1-12 Shinzaike-honcho, Himeji, Hyogo 670-0092, Japan

^{††} Graduate School of Human Science and Environment, University of Hyogo
1-1-12 Shinzaike-honcho, Himeji, Hyogo 670-0092, Japan

^{†††} School of Human Science and Environment, University of Hyogo
1-1-12 Shinzaike-honcho, Himeji, Hyogo 670-0092, Japan

E-mail: [†]{na03z056,nd06c019,nd05r011}@stshse.u-hyogo.ac.jp, ^{††}sumiya@shse.u-hyogo.ac.jp

Abstract Presentation with a heterogeneous media such as videos and slides has come frequently to be done now. Moreover, a variety of retrieval methods are proposed to those stored media. We proposes the retrieval method based on the characteristic of those each media data. In our proposed method, the most suitable interval for query is retrieved by making the use media of the stored presentation at attention video and slide, and considering which media query keyword that the user input appears. As a result, the interval that more efficient interval for the viewer only who doesn't depend on the presence of query keyword can do can be extracted.

Key words Infomation integration, Personalization, Information retrieval, Spacio temporal DB

1. はじめに

近年，プレゼンテーションをビデオ録画する環境が整ってきた [1]．それにより，会議や講義を当日欠席した者が後日視聴することや，発表者や聞き手が内容を知った上で再度視聴することができるようになった．また，それらのデータを蓄積したアーカイブコンテンツを検索することで視聴したいプレゼンテーションを抽出することも行われている．

本研究では，プレゼンテーション視聴システムを利用する際

に，ユーザの意図を入力することで 1 つのプレゼンテーションから部分的なシーンを抽出する方式を提案する．本方式の特徴は 2 点あり，1 点はユーザの意図のタイプを判定することでより適切な区間抽出を行うこともう 1 点はメディアの特性を考慮して区間の抽出と出力を行うことである．ここでいう意図とはユーザの質問であり，メタデータに含まれるキーワードを指している．

本稿で扱う異種メディアとは，プレゼンテーションに用いられるスライドとビデオを指す．我々は 2 つのメディアごとに存

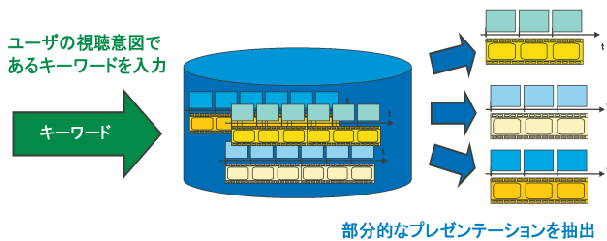


図1 概念図

在する各メタデータには、スライドにのみ存在するキーワード、ビデオ中の発話にのみ存在するキーワードがあることに着目し、視聴者の入力に対して1つのプレゼンテーションから適切な出力区間を判定する方式を提案する(図1)。

以下、2節では本研究の動機と関連研究について説明する。3節では、視聴者の質問により抽出される部分区間の結合方法の概論を述べる。4節では、区間結合方法をプレゼンテーションコンテンツに反映した手法を述べる。5節では、プレゼンテーションコンテンツにおけるユーザの意図の判定方法と出力区間の決定方法、そして出力方式について述べる。6節では、プロトタイプシステムについてと、本手法の評価方法とその考察を述べる。7節ではまとめを述べる。

2. 本研究のアプローチと関連研究

2.1 本研究のアプローチ

我々は、ユーザが区間検出のために入力したクエリキーワードがどのメディアのメタデータとして含まれているかによって、同じクエリキーワードであったとしても出力すべき区間が異なってくるのではないかと考えた。そこで、クエリキーワードの有無を調べるメディアの対象を、プレゼンテーションを録画した際に存在するスライドとビデオという2つのメディアとして議論する。

それらのメディアの特徴を考慮することで、元データの中から視聴者の入力に対して最適であると判定できる区間を出力することが本研究の目的である。例えば、発表者が新商品のゲーム機器についてプレゼンテーションを行う際、スライド上に記載されている「特長」というキーワードを口述してそのスライドを説明するだけでなく、スライドにおいて記載されていない「特長」というキーワードを使って提示しているスライドを説明することや、突発的に挿入された発表者の私的な余談の中に「特長」が含まれている場合がある。同じ「特長」というキーワードでも、そのそれぞれで重要度は異なるはずである。これらの重要度を考慮し、視聴者の効率的な視聴を実現する。

本方式では、ユーザは過去のプレゼンテーションビデオ、およびビデオと同期したスライドに対し、自らが興味のあるキーワードをクエリとしてシステムに入力する。出力として、単にキーワードが存在する区間が提示されるのではなく、選出区間の中でよりキーワードの重要度が高いものが検索される。

システムは、まず入力されたクエリキーワードを用いてキーワードのタイプ判定を行い、元データから出力候補となりうる区間を抽出する。次に、抽出された出力候補区間の中で判定

されたクエリキーワードによる重要度の算出を行う。更に、メディアに基づく重要度をそれぞれ統合する演算を行い、抽出区間ごとにクエリキーワードの重要度を算出し、検索結果のランキングを行う。

2.2 関連研究

近年、映像や音声などの複数のメディアが統合されてきたコンテンツを蓄積、検索する研究が数多く提案されている。例えば、ビデオ区間集合から新しいビデオ区間を生成するグルー結合[6][9]や、実況者の発話や記事から映像のメタデータを生成してビデオ区間を検索する手法[8]がある。

対象が映像だけでなく、スライド情報を用いたプレゼンテーションコンテンツについて、岡本ら[4]は、スライドと映像の同期情報やキーワードの適合性から新しいメタデータの生成手法[2]を提案している。資料と映像の疎結合するために、メタデータの抽出方法と格納方法[3]や、音声情報やポインタ照射情報[5]を用いた手法など数多くの研究が進んでいる。また、プレゼンテーション中だけでなく質疑応答のシーンを用いたアーカイブコンテンツの作成方式[7]等も提案されている。

一方で、抽出した部分的な区間とユーザの検索意図との適合性については十分な検索が行われているとはいえない。そこで我々は、既存のスライド資料と映像そのものを2つのメタデータとして考え、資料と映像の同期が目的ではなくユーザの意図に適合した部分的な区間の抽出を目指す。また、本システムでは既に存在するメタデータを新しく分類し直して検索を行っている点で検索対象が異なっているといえる。

3. 同種・異種メディアの統合

3.1 メタデータとその利用

本研究でメディアの統合を行うために7つのメタデータを使用する。メタデータは、同種メディアの統合に用いるメタデータと異種メディアを統合する時に用いるメタデータを定義する。また、異種メディア統合に用いるメタデータは同種メディアの統合に用いるメタデータを包含する(図2)。

メタデータの基準になっているものはユーザがコンテンツを視聴する目的となるクエリキーワードである。ユーザの目的であるクエリキーワードを用いることで発表側の意図や発表内容だけでなく、視聴者側の意図を考慮したメディアの統合を行う。これらメタデータを使用してユーザの目的に応じたメディアの統合を実現する。

また、メディア統合を行う際にメタデータだけではなく各メディアの固有のルールを用いる必要がある。固有のルールとは、統合対象のメディアの特性を反映した個々のルールであり、例えば統合対象が映像であった場合、映像に存在する時間情報を反映したルールなどである。

3.2 メディアの統合

3.2.1 同種メディアの統合

視聴者の目的が全て同種のメディア内に存在している時、同種メディアにおいて統合を行う。同種メディアの統合には以下の3つのメタデータを主に用いる。

- ユーザの目的オブジェクトが存在する単位区間数

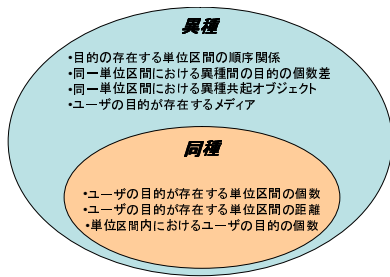


図2 メディア統合における使用メタデータ

- ユーザの目的オブジェクトが存在する単位区間同士の距離

- 各単位区間におけるユーザの目的オブジェクトの個数

3.2.2 異種メディアの統合

異なるメディア間で視聴者のクエリが存在している時、異種メディアの統合を行う。異種メディアの統合には3つのメタデータに加え、4つのメタデータを用いる。これらを反映して統合を行う。

- ユーザの目的オブジェクトが存在するシーン数
- 目的オブジェクトが存在する単位位置の順序関係
- 同一単位において異種間での目的オブジェクトの個数差
- 他方のメディアにおいて同一単位内で目的オブジェクトと共起するオブジェクトの存在有無

- ユーザの目的が存在するメディアの種類

異種メディアにおいては目的オブジェクトの単位位置や個数だけでなく、存在する単位同士やメディアの関係を考慮している。これにより、別のメディアを用いることで各メディアに適合した区間が抽出される。

これらのメタデータは各メディアの固有の特性を反映する。さらに、クエリが存在するメディアの種類により各固有のメタデータを反映する。固有のメタデータを反映するのは、存在するメディアによってコンテンツ内でのユーザが目的とするオブジェクトの種類が変化しているためである。

4. プレゼンテーションコンテンツ

4.1 プレゼンテーションコンテンツのメタデータ

本節では、3節で述べたメディア統合の手法をスライドとビデオを用いたプレゼンテーションコンテンツを対象に説明する。ここではメディアの単位をシーンとし、1シーンとは各スライド、また各スライドに対応したビデオ区間を指すものとする。

プレゼンテーションコンテンツにおいては、視聴者の目的オブジェクトを視聴者が入力するクエリキーワードとし、視聴者はプレゼンテーションコンテンツを視聴しながら、好きなキーワードを入力できる。よって前節で述べた同種メディアの統合に利用するメタデータは以下の三つである。

- クエリキーワードが存在するメディアのシーン数
- クエリキーワードが存在するメディアのシーン間の距離
- 各メディアの各シーンにおけるクエリキーワード数

また、異種メディアの統合に利用するメタデータは、前節より以下ようになる。さらに、クエリが存在するメディアの種類により各固有のメタデータを反映する。

- ユーザのクエリキーワードが存在するシーンの個数
- クエリキーワードが存在するシーンの順序関係
- 同一シーンにおいて異種間でのクエリキーワードの個数差
- 他方のメディアにおいて同一シーン内でクエリキーワードと共起するキーワードの存在有無
- クエリキーワードが存在するメディアの種類

ここではメディアはスライドとビデオの2つである。クエリがビデオに存在するとき、メディア単位数はより少ない場合、メディア単位間の距離はより長い場合とより短い場合、そして各単位におけるクエリキーワード数は多い場合に重要と判定される傾向がある。同様に、クエリがスライドに存在するとき、メディア単位数はより多い場合とより少ない場合、メディア単位間の距離はより短い場合、そして各単位におけるクエリキーワード数はより少ない場合に重要と判定されることがある。

クエリがスライドとビデオの両方に存在するとき、メディア単位数はより多い場合とより少ない場合に重要と判定され、各単位におけるクエリキーワード数はメディア間での差の個数を利用する。これはビデオに比べスライドのメタデータがより規則的に発生しやすく、単位の発生数に依存しにくい特性を反映している。

4.2 区間結合

プレゼンテーションコンテンツを視聴する時、最初から最後まで全体を視聴したり部分的に視聴したりすることがある。本研究では視聴者が入力したクエリキーワードを用いて検索を行い、クエリが存在するシーンを区間とし、それらを結合して抽出する。また、抽出された結合区間は視聴者の入力するクエリキーワードに対する出力の候補区間となる。

結合にあたっての手法は基本的にグルー結合 [6] を用いるが、クエリが存在するシーンを区間としているため、単一区間を抽出することがある。つまり、クエリキーワードが存在するシーンを区間として結合し、出力の候補区間を複数生成する。そしてクエリキーワードの性質判定をし、クエリに対するこれらの候補区間の各適合度を算出し出力を決定する。

5. プレゼンテーションコンテンツの統合方式

本稿において、プレゼンテーションコンテンツの統合を行うために、クエリキーワードの性質を決定するタイプ判定と、結合した複数の候補区間から最も適合した区間の判定という二つの判定を行う。また、最適な区間を決定するために各結合区間の重要度の算出について述べる。以下、5.1節において二つの判定について述べ、5.2節で各結合区間の適合度算出について述べる。

5.1 質問キーワードのタイプ判定

クエリのタイプを判定するために、プレゼンテーションにおいてキーワードの性質を設定する必要がある。そこで、クエリキーワードの型を「メインピック (Main-Topic)」「サブピック

表1 クエリキーワードのタイプ判定

	シーン数	シーン間の距離	クエリの個数	シーンの順序関係	クエリの個数の差	共起キーワード	メディアの種類
メインピック	多い	近い	-	-	小さい	-	Slide or Video
サブピック	少ない(共起したもの)	-	-	-	大きい	-	Slide or Video
バックグラウンド	多い	-	少ない	×	×	×	Slide
例	-	-	-	-	-	有	Video
参照	少ない	近い	-	有	-	-	Slide or Video
付帯トピック	少ない	遠い	-	×	×	×	Video
判定不能	-	-	-	-	-	-	-

ク (Sub-Topic)」「文脈参照 (Reference)」「例 (Example)」「バックグラウンド (Background)」「付帯トピック (Extra-topic)」の6個に設定する。また、上記6つのタイプに判定されなかったキーワードは「判定不能 (Irregular)」と判定する。以下にそれぞれの定義を述べる。

メインピック 発表内容の主旨

サブピック 発表内容の部分的なポイント

文脈参照 表示中のスライドに関連した別シーン

例 発表者が口頭でスライドに記載されていない言葉で行う説明

バックグラウンド 発表内容の前提

付帯トピック 発表内容とは関係のない余談

これら6種類のタイプは各コンテンツ内においての出現傾向により二つに大別される。一つは始めから終わりまで全体的に出現している場合、もう一つは局所的に出現している場合である。以下に出現傾向別に分類したクエリのタイプと性質を述べる。

<全体的に出現するクエリのタイプ>

バックグラウンド スライドには記載されるが主旨の前提であるためビデオには出現しない。

付帯トピック 発表内容とは関係なく挿入されるためビデオにのみ出現する。

例 発表者がより詳細に行う説明のためスライドに出現しない言葉を用いた文章が挿入される。

メインピック 同じシーンにおいてスライドとビデオにそれぞれ出現する。またそのシーン数が多く、及び同シーンにおいてメディア間では出現するクエリの個数の差がほぼ発生しない。

<局所的に出現するクエリのタイプ>

文脈参照 1シーンにおいては1メディアにだけ出現する。また、一つのシーンにおいて一時的に別のスライドに切り替わり発生する。

メインピック 同じシーンにおいてスライドとビデオにそれぞれ出現する。またそのシーン数多く、さらに同シーンにおいてメディア間では出現するクエリの個数に差が発生しうる。

サブピック 同じシーンにおいてスライドとビデオにそれぞれ出現するが、一方のメディアにのみ出現するシーンが多く出現する。及び同シーンにおいてメディア間では出現するクエリの個数に大きな差が発生しうる。

これらのタイプは前に述べたメタデータの組み合わせにより

判定ができる(表2)。例えば入力されたクエリキーワード a がコンテンツ内でスライドにだけ存在、そのシーン数も多く、そして1シーン中の a の個数が少ない時、クエリキーワード a は「バックグラウンド」と判定される。

5.2 区間重要度の算出

判定されたクエリキーワードのタイプから、そのプレゼンテーションコンテンツ内でのクエリキーワードによる区間の重要度を算出する。クエリキーワードのタイプは1つのプレゼンテーション中での出現傾向により判定されるため、それらの要素を用いて区間の重要度を算出する。また、定義した6種類のクエリキーワードのタイプに加え、クエリキーワードが上記6種類のいずれにも該当しなかった場合の「判定不能 (Irregular)」の重要度も算出方法を定義する。

まず、クエリキーワード a が「メインピック」と判定された場合の重要度は、以下の式で算出する。

$$Main(a) = \frac{S_{num}(a)}{\frac{\sum_{i=1}^{n-1} (dist(a_i, a_{i+1}))}{n-1}} \quad (1)$$

S_{num} はプレゼンテーション中にクエリキーワード a が存在するシーン数であり、 $dist(a_i, a_{i+1})$ はクエリキーワード a が存在するシーン同士の距離を算出している。また、 $differ(a)$ は各シーンでスライドとビデオを比較したときのクエリキーワード a の個数差を表す。

次に、クエリキーワード a が「サブピック」と判定された場合の重要度は、以下の式で算出する。

$$Sub(a) = \frac{differ(a)}{S_{num}(a_S \wedge a_V) + 1} \quad (2)$$

$S_{num}(a_S \wedge a_V)$ は1つのプレゼンテーション中でクエリキーワード a がスライドとビデオの両方に出現するシーン数を表す。

クエリキーワード a が「文脈参照」と判定された場合、以下の式で算出する。

$$Reference(a) = \frac{1}{S_{num}(a) \times dist(a_{Sstart}, a_{Vlast}) + 1} \quad (3)$$

$$Reference(a) = \frac{1}{S_{num}(a) \times dist(a_{Slast}, a_{Vstart}) + 1} \quad (4)$$

2パターンの式が存在するのは、「文脈参照」の性質として、クエリキーワード a が一方のメディアに出現後にもう一方のメディアに出現するためである。これらのシーン間の距離を表したものが $dist(a_{Sstart}, a_{Vlast})$ と $dist(a_{Slast}, a_{Vstart})$ である。

「例」と判定された場合、以下の式で算出する。

$$Exam(a) = \sum_{i=1}^n (co - oc_{num}(a_V, K_{Si})) \quad (5)$$

K_{Si} はクエリキーワード a と同シーンで共起するキーワードを表しており、 $co - oc_{num}(a_V, K_{Si})$ はそれらの共起したシーン数を表す。

「バックグラウンド」と判定された場合、以下の式で算出する。

$$BG(a) = \frac{S_{num}(a)}{Key_{num}(a) + 1} \quad (6)$$

$Key_{num}(a)$ は1つのプレゼンテーション中でのクエリキーワード a の出現数である。

そしてクエリキーワード a が「付帯トピック」と判定された場合、

$$Ext(a) = \frac{\sum_{i=1}^{n-1} (dist(a_i, a_{i+1}) + dist(1, a_1) + dist(a_m, n))}{S_{num}(a) + 1} \quad (7)$$

と算出される。 $dist(a_i, a_{i+1})$ はクエリキーワード a が存在する隣り合うシーン間の距離を表し、 $dist(1, a_1)$ はプレゼンテーションの始まりからクエリキーワード a が初めて出現した距離を表している。また、 $dist(a_m, n)$ はクエリキーワード a が一番後ろに出現したシーンからプレゼンテーションの最後のシーンまでの距離を表している。

最後に、クエリキーワード a が上記6種類のいずれにも判定されなかった「判定不能」の場合の重要度は、以下のように算出される。

$$I(a) = \frac{Key_{num}(a)}{dist(a_i, a_j)} \quad (8)$$

これらは1つのプレゼンテーション中での出現傾向から算出した単一のクエリキーワードによる区間の重要度であり、抽出された各出力候補区間にクエリキーワードが単一メディアのみ出現していた場合、その出力候補区間の重要度となる。

1つの出力候補区間に複数の種類のクエリキーワードが存在した場合、出力候補区間の重要度は各クエリキーワードの重要度同士の要素として算出する。それらを示す算出方法を表2にまとめる。例えば、視聴者が入力した a, b というキーワードがそれぞれ「メイントピック」と「バックグラウンド」と判定されて抽出された出力候補区間の重要度は、 $Main(a)$ と $BG(b)$ を式(1), (3)から算出し、表2から $Main(a) + BG(b)$ が出力候補区間の重要度となる。このように、抽出された全ての出力候補区間の重要度をそれぞれ算出することにより、重要度の値で出力候補ランキングを生成できる。

抽出された全ての出力候補区間長に対して全区間長に対する最大区間長のフィルタを設ける。ここでは、全体の1/2を上限とすることとする。

5.3 プレゼンテーションコンテンツの視聴

出力区間には、各クエリキーワードに該当し、重要度を持った部分区間が存在する。入力されたクエリキーワードが単数の場合、それが出現する区間がクエリに該当した区間であるといえる。表3に詳細を示す。

表3 出力メディアの決定基準

タイプ	出力メディア
メイントピック	Slide and Video
サブトピック	Slide and Video
バックグラウンド	Slide
例	Slide and Video
文脈参照	Slide and Video
付帯トピック	Video
判定不能	Slide and Video

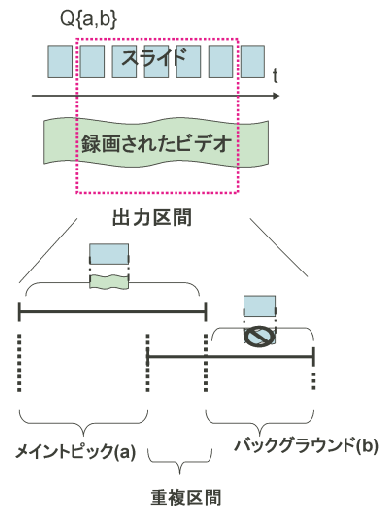


図3 各重要度を用いた視聴形態の決定

一方、クエリキーワードが複数存在する場合、結合された区間は複数の重要度を持った区間から構成されており、部分的に重要度が重複する区間が発生しうる。図3のように2つクエリが入力され、それぞれが a, b が「メイントピック」と「バックグラウンド」と判定され、出力区間が重複して発生した場合、各該当区間の重要度を比較してより大きいタイプの出力形態が重複区間の出力形態とする(図3)。

6. 評価

6.1 プロトタイプシステム

プロトタイプシステムは、クエリキーワードの判定部と重要度算出部に分けられる。これらを Visual Studio.NET のC#により構築する。これはビデオとスライドのメタデータから視聴者が入力するクエリキーワードを検索し、出力候補区間を生成する。そして、各区間の重要度が最大の区間を出力区間と判定し、その重要度を提示するものである。ビデオからの単語抽出には茶筌^(注1)による形態素解析を用いて抽出する。

図4で用いたプレゼンテーションは、スライドとビデオの二つのメディアを使用した講演ビデオである。視聴者がビデオを閲覧している中に気になったキーワードを入力すると同じ講演ビデオの全区間から最適だと判定できる区間をシステムが自動的に検索し、提示することができる。また、ランキングを行った最上位ではない抽出区間も提示し閲覧できるようにする。

(注1): <http://chasen.naist.jp/hiki/ChaSen/>

表2 区間結合における重要度の算出方法

	メイントピック	サブトピック	バックグラウンド	例	参照	付帯トピック	判定不能
メイントピック	+	+	+	+	+	×	×
サブトピック	+	+	×	+	+	+	×
バックグラウンド	+	×	+	×	×	×	×
例	+	+	×	+	+	+	×
文脈参照	+	+	×	+	+	×	×
付帯トピック	×	+	×	+	×	+	×
判定不能	×	×	×	×	×	×	+

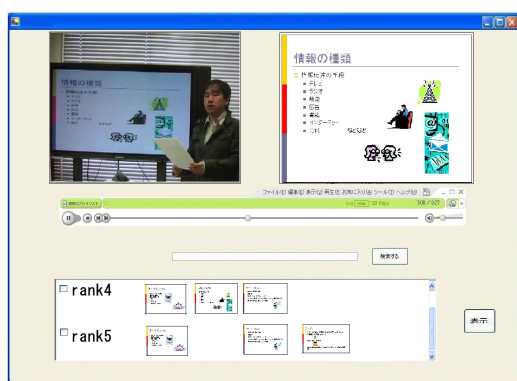


図4 インターフェース例

プロトタイプシステムの画面の上部は視聴区域であり，中心部に検索ツールを設置した．そして検索した抽出区間をランキング提示し，右側には該当区間のスライドを提示している．

6.2 実験

考案した手法に対し評価実験を行った．実験では，入力したクエリキーワードに対して出力区間と出力が適合しているかどうかを判定するものである．

実験に使用したデータは，実際の有効性を判定するため，実際に行われたプレゼンテーションコンテンツの3組を用いた．用意したデータに対して，同時に複数のキーワードが入力された場合と単数のキーワードが入力された場合との2つのパターンを考慮するため，1つのデータに対して2個のキーワードを同時に指定した場合と1つのキーワードを指定した場合をそれぞれ3パターンずつ行った．また，各データごとでキーワードは6個ずつ用いており，計18回の実験を行った．発表データ1はスライド17枚とそれを録画した14分19秒のビデオ，発表データ2はスライド26枚と17分32秒のビデオ，発表データ3はスライド25枚と14分52秒のビデオを用いた．

6.2.1 発表データ1

単数キーワードのパターンは“P2P”，“検索”，“参加”の3つであり，複数キーワードのパターンは“P2P モバイル”，“検索 ユーザ”，“参加 要求”の3つである．“P2P モバイル”の場合の抽出結果とその区間図を図5に示す．使用したデータの17シーンの内“モバイル”が出現したのはスライド番号の2, 5の計2シーンであり，2番目はスライドとビデオに出現し5番目はスライドのみの出現であった．“P2P”はスライド番号が1,4,5,6,17の計5シーンに登場しており17番目のシーン以

外は両方のメディアに出現していた．またキーワード“モバイル”自体の出現個数は各メディアの各シーンに1つずつの出現であったが，“P2P”はスライド番号の4と5の各シーンにおいてメディア間での個数差がみられた．

このデータ1を用いて複数キーワード“モバイル”，“P2P”を同時に入力した際の算出結果を説明する．キーワードを入力した結果， i_1 から i_4 の4区間の出力候補区間が抽出され，重要度が0.22で最大となった i_1 が出力区間となった．但し，クエリキーワードの出現する箇所だけではなく出現する区間を生成するため，生成された区間に含まれるクエリキーワードは全てカウントされるとする．

キーワードタイプの判定は，“モバイル”が「文脈参照」，“P2P”が「判定不能」として判定された．これは“モバイル”がプレゼンテーション17シーン中の2シーンだけ出現していることやスライド番号が5のシーンはスライドのみの出現であったことと，“P2P”は出現した計5シーン中4シーンが序盤に存在し，ビデオ中での出現個数が多かったためである．これは“モバイル”の出現頻度が低いことと，“P2P”の偏在的な頻出性が主として反映されていると考えられる．しかし，“P2P”においてはメディア間での出現個数の差が僅かながら発生してため「判定不能」と判定された．

“モバイル”の重要度は0.11の1区間が算出されるが，“P2P”の重要度は時区間のフィルタリングにより4区間算出され，その中から区間重要度を算出する際に全て2.0と判定される．それにより，区間重要度が全て同値として算出され最短区間1が出力区間となる．

この区間は“モバイル”の「文脈参照」区間が“P2P”の「判定不能」区間に含まれている区間であるため，「文脈参照」の区間が重複区間であるといえる．各重要度を比較した際，「文脈参照」の重要度は0.11でありこれは「判定不能」のどの重要度よりも小さい値である．よって「判定不能」の出力方式で出力するため，出力区間のメディアはスライドとビデオの同時再生となる．

6.2.2 発表データ2

単数キーワード“結合”を入力した際の算出結果を説明する．キーワードを入力した結果， i_1 から i_{10} の10区間の出力候補区間が抽出され，重要度が5.0で最大となった i_1 が出力区間となった．

キーワードタイプの判定で“結合”は，全データの内部的に集中して出現しており，メディア間での個数差が大きい

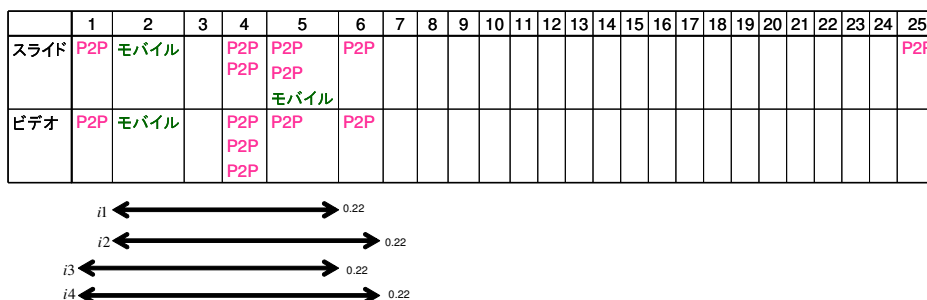


図5 発表データ1に複数キーワード {P2P, モバイル} を入力

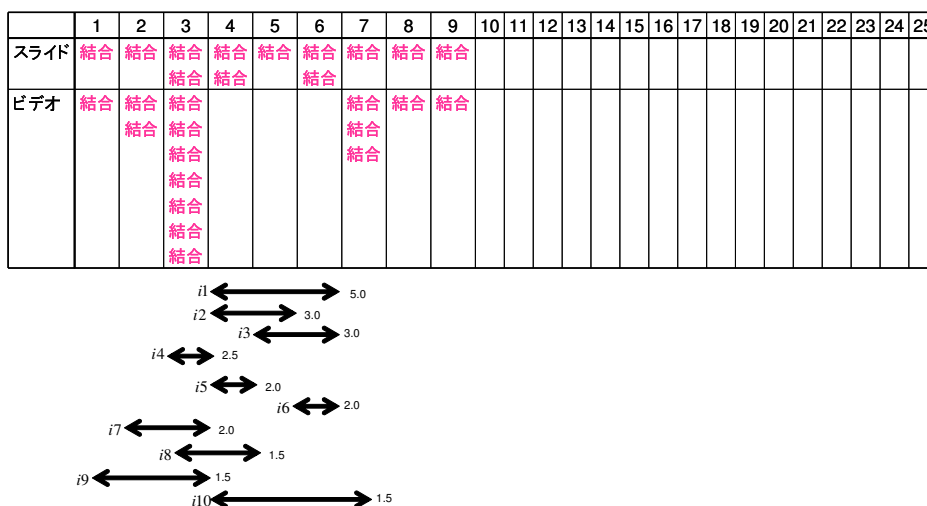


図6 発表データ3に単数キーワード {結合} を入力

「サブピック」として判定された。「結合」の重要度は時区間のフィルタリングにより全45区間算出される。その中から区間重要度を算出する際に全て2.0と判定される。それにより、区間重要度が5.0, 3.0, 3.0, 2.5, 2.0, 2.0, 2.0, 1.5, 1.5, 1.5と算出される。よって、重要度が5.0で最大となった i_1 が出力区間となる。また、これは「サブピック」の出力方式で出力するため、出力区間のメディアはスライドとビデオの同時再生となる。

6.3 考察

6.3.1 発表データ1

“P2P”と“モバイル”を同時に入力した場合では、出力区間としては妥当な結果が出たが、キーワードのタイプ判定としては再考すべき点を発見した。適切だと判定できたのは、区間長が最短であり意味的なまとまりをもった区間が出力として決定されたことであり、不適切だと考えられるのは“P2P”のキーワードのタイプが「判定不能」と判定したことである。

重要度の算出結果を見てみると、「文脈参照」としての“モバイル”はスライド番号2~5を1つのシーンとして考えるため、「判定不能」である“P2P”が存在するシーンが多く含まれた区間の重要度が最高になる。しかしながら、今回“P2P”の出現個数和が、区間が異なった場合も、区間長が異なるものの各重要度が一致する現象が発生した。これは、キーワードそのものだけでなく、出現状況を考慮した各シーンでの意味づけを考慮する必要が有ると考えられる。実データを視聴した結果、図5

の出力区間が“モバイル”を引用しているシーンと“P2P”の頻出シーンを最も含んでいることが分かった。

“モバイル”は実データを視聴すると、スライド番号2での出現をスライド番号5で引用していたため、「文脈参照」の判定は妥当であると考えられる。

しかしながら、キーワード“P2P”はデータ中である程度出現しており、またタイトルの一部にも登場しているにも関わらず「判定不能」として判定された。これは、出現シーン数とクエリの個数が閾値以上存在しなかったためであり、プレゼンテーション中でタイトルの一部として出現しているなどの位置づけを考慮すると、この判定は適切でないことが考えられる。

該当シーン数や出現個数に依存してランク付けが行われている面があるため、同じ“P2P”でもスライド番号1でのタイトルとその読み上げシーンが含まれる区間と、代わりに詳細な説明の終焉が含まれる区間が同等に扱われてしまっている。よって判定条件に偏りが存在していると判断されるため、再考すべきと思われる。

6.3.2 発表データ2

キーワード“結合”を入力すると、「サブピック」であると判定された。これはタイトルの一部にも出現しているキーワードであり、出現頻度や個数を考慮しても「サブピック」の判定は妥当であると思われる。

しかし、出力区間の判定は不適切と判断できる(図6)。その理由は、出力該当区間はスライドが存在するものの、急遽発表

者が早送りを行って口頭説明が全く存在しなかった区間だったためである。つまり、説明が全くされなかったシーンとクエリキーワードが出現しなかっただけのシーンが同等に扱われているということが考えられる。他の出力候補区間においても、それぞれで意味をもったまとまりが存在するが、出力として適切であるとはいい難いと思われる。

今回の実験で、2つのシーンに該当し、後者が前者の説明をしているキーワードを「文脈参照」と判定するため、「文脈参照」の判定は妥当であったと考えられる。また、多くの場合で、プレゼンテーションの主旨を示すキーワードを「メインピック」であると判定するなど、タイプ判定は概ね有効であると考えられる。しかしながら、意図的に発話者が発したキーワードでも、出現回数が少ないと、そのキーワードを「付帯ピック」と判定してしまうことも分かった。

重要度の判定については、タイプ判定を行うことで区間を抽出しているため、各区間は意味的なまとまりが見られた。しかしながら、ランキングの上位に意味の無い区間が含まれることもあった。また、時間の都合上、省略されたスライドなどを考慮できないため、区間抽出がプレゼンテーションの精度に左右することがあった。

出力メディアは出力区間がスライドとビデオの同時再生と決定したが、実験で用いたケースでは出力区間がクエリキーワードと関係が低いと思われるシーンを多数含んでおり、全てを二つのメディアで出力する有効性が低いのではないかと考えられる。よって出力するメディアの決定条件を再考する必要がある。

6.4 ま と め

本稿において、プレゼンテーションを視聴システムのための異種メディア統合方式を提案した。プレゼンテーションのコンテンツに対して、ユーザの入力するキーワードで以って区間抽出を行うが、入力されたクエリキーワードの出現傾向からまずタイプ判定を行い、そのタイプで抽出区間の重要度が決定されて出力区間を決定する方式である。

さらに、我々は、実データを用いて本方式の評価実験を行った。実験では、クエリのタイプ判定は概ね有効であるが、区間抽出はそのプレゼンテーションの精度に左右されることが分かった。

今後の課題としては、キーワードのさらなる詳細な種別化、もしくはシーン群の種別化が必要であるといえる。また、ノイズが含まれる対策として時区間のフィルタリングも設置したが、そのフィルタリングも種別化に伴い改良する余地があると思われる。

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)、および西兵庫信用金庫助成金によるものです。ここに記して謝意を表すものとします。

文 献

[1] MPMeister. 株式会社リコー. <http://www.ricoh.co.jp/mpmeister/>.

- [2] 村木太一, 吉田誠, 小林隆志, 直井聡, 横田治夫. メタデータによる講演資料と動画の統合と検索. DBWeb2002, 情報処理学会, pp.97-104,2002.
- [3] 小林隆志, 村木太一, 直井聡, 横田治夫. 統合プレゼンテーションコンテンツ蓄積検索システムの試作. 電子情報通信学会論文誌, 電子情報通信学会, Vol.J88-D-I, No.3, pp.715-726,2005.
- [4] 岡本拓明, 仲野亘, 小林隆志, 直井聡, 横田治夫. プレゼンテーション蓄積検索システムにおける講義・講演音声情報を利用した適合度の改善. 第17回データ工学ワークショップ (DEWS2006), 6c-01.
- [5] 仲野亘, 小林隆志, 勝山裕, 直井聡, 横田治夫. 講演シーン検索における検索語出現状況に基づくレーザーポインタ情報のフィルタリング. 日本データベース学会論文誌 (DBSJ Letters), Vol.5, No.2.
- [6] 西尾章治郎, 田中克己, 上原邦昭, 有木康雄, 加藤俊一, 河野浩之. 情報の構造化と検索. 岩波書店, 2000.
- [7] 戈指夷, 角谷和俊. 遠隔会議システムにおける資料操作ログに基づくアーカイブコンテンツ作成支援方式. 第17回データ工学ワークショップ (DEWS2006), 6C-i5.
- [8] 吹野直紀, 馬強, 角谷和俊, 田中克己. ニュース記事を利用したサッカー要約映像の生成. 第14回データ工学ワークショップ (DEWS2003), 8-P-03.
- [9] プラダンスジツト, 田島敬史, 田中克己. ビデオデータ検索のための区間グルー操作とフィルタリング. 情報処理学会論文誌, Vol.40 No.SIG3(TOD1), 1999.
- [10] DBSJ アーカイブ. , 日本データベース学会. <http://www.dbsj.org/Japanese/Archives/archivesIndex.html>.