

論文 / 著書情報
Article / Book Information

論題(和文)	講義講演ビデオからダイジェスト自動作成のための重要シーン抽出手法の評価
Title(English)	Evaluation of Methods to Extract Important Scenes for Automatic Digest Generation from a Presentation Video
著者(和文)	レー ヒェウハン, ティティポーン ルートラットデーチャークン, 渡部 徹太郎, 横田 治夫
Authors(English)	Le HIEU HANH, LERTRUSDACHAKUL Thitiporn, Tetsutaro WATANABE, HARUO YOKOTA
掲載誌(和文)	データ工学ワークショップ2008 論文集
Citation(English)	Proc of DEWS2008
Vol, no, pages	, , E4-1
発行日 / Pub. date	2008, 3
URL	http://search.ieice.org/
権利情報 / Copyright	本著作物の著作権は電子情報通信学会に帰属します。 Copyright (c) 2008 Institute of Electronics, Information and Communication Engineers.

講義講演ビデオからダイジェスト自動作成のための 重要シーン抽出手法の評価

レー ヒェウハン[†] ルートラットデーチャークン ティティポーン^{††} 渡部 徹太郎^{†††}
横田 治夫^{†††,†††}

[†] 東京工業大学 工学部情報工学科 〒 152-8552 東京都目黒区大岡山 2-12-1

^{††} (株) リコー ソフトウェア研究開発本部アプリケーション研究所 〒 112-0002 東京都文京区小石川 1-1-17
とみん日生春日町ビル

^{†††} 東京工業大学大学院 情報理工学研究科 計算工学専攻 〒 152-8552 東京都目黒区大岡山 2-12-1

^{††††} 東京工業大学 学術国際情報センター 〒 152-8550 東京都目黒区大岡山 2-12-1

E-mail: †{hanhllh,tetsu}@de.cs.titech.ac.jp, ††thitiporn.lertrusdachakul@nts.rioh.co.jp,

†††yokota@cs.titech.ac.jp

あらまし E-ラーニングなどにおいて、講義講演の様子を録画したビデオを利用する機会が増えている。十分な時間的余裕がない場合等、講義講演のビデオのダイジェストのみを見たい場合もある。本稿では、動画とスライドからコンテンツを作成するシステムである MPMeister で収録された講義講演の情報から、重要シーンを抽出しダイジェストを自動的に作成する手法を提案する。重要シーン抽出においては、講義講演で使用されるスライド中に出現する単語の出現頻度、特定性、提示時間、シーンの順序を考慮した複数の手法を検討する。また、抽出された重要シーンと被験者により選択された正解セットを比較することで、それぞれの重要シーン抽出手法の結果を評価する。
キーワード e-ラーニング、情報抽出、ビデオダイジェスト

Evaluation of Methods to Extract Important Scenes for Automatic Digest Generation from a Presentation Video

Le HIEU HANH[†], Thitiporn LERTRUSDACHAKUL^{††}, Tetsutaro WATANABE^{†††}, and
Haruo YOKOTA^{††††,†††}

[†] Department of Computer Science, Faculty of Engineering, Tokyo Institute of Technology Ookayama 2-12-1,
Meguro-ku, Tokyo, 152-8552 Japan

^{††} Application Lab, Software R&D Group, Ricoh Company, Ltd. Tomin-Nissei-Kasugacho Bldg.Koishikawa
1-1-17, Bunkyo-ku, Tokyo, 112-0002 Japan

^{†††} Department of Computer Science, Graduate School of Information Science and Engineering, Tokyo
Institute of Technology Ookayama 2-12-1, Meguro-ku, Tokyo, 152-8552 Japan

^{††††} Global Scientific Information and Computing Center, Tokyo Institute of Technology Ookayama 2-12-1,
Meguro-ku, Tokyo, 152-8550 Japan

E-mail: †{hanhllh,tetsu}@de.cs.titech.ac.jp, ††thitiporn.lertrusdachakul@nts.rioh.co.jp,

†††yokota@cs.titech.ac.jp

Abstract In the recent years, the use of presentation video in E-learning is increasing. On the user's point of view, this often requires previewing the digest of such video before watching the original one. This paper proposes a number of methods of extracting important scenes for automatic digest generation digest from the presentation video which is recorded by MPMeister II - the tool for Multimedia Web Contents. The important scenes extractions are based on several factors, such as word frequency, specificity, duration and orders of the slides. Finally, the results of each methods are evaluated by comparing with the answer sets selected by testers.

Key words e-learning, information extraction, video digest

講義の録画

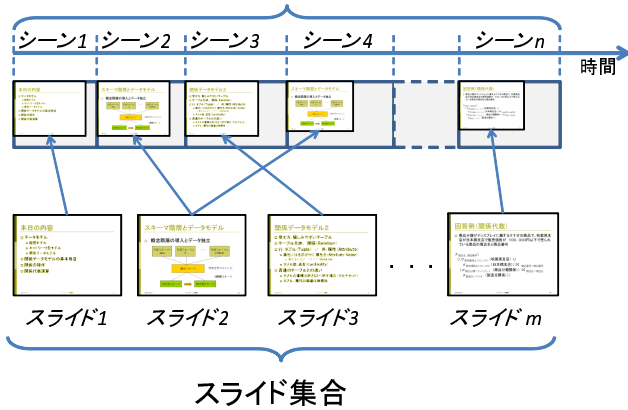


図1 シーンとスライドの対応

1. まえがき

近年、プレゼンテーションソフトウェアを用いた講義講演を録画し、E-ラーニングなどに用いる機会が増えてきている。(株)リコーによって開発されたMPMeister [1]のようなソフトウェアによって容易にコンテンツ作成できるようになった影響が大きい。

これまで我々は、そのような講義講演コンテンツから、利用者が示す検索キーワードに合致するシーンを探し出すために、UPRISE [2]を提案し、その効果を示してきた。シーンとはスライドの切り替えによる動画の部分である(図1参照)。しかし、講義講演コンテンツの場合、利用者は検索キーワードに関して適切な知識を持ち合わせているとは限らず、検索が有効でない場合もある。また、当該講義講演の概要を短時間で把握したいという要求も高い。従来、動画要約の研究も多くされているが、講義講演コンテンツの特徴と異なるため、それらの手法を本研究にそのまま適用することができない。

そこで、本稿ではMPMeisterで録画された講義講演コンテンツから、講義講演のトピックを良く表すシーン(以降重要シーンと呼ぶ)を抽出し、ダイジェストを自動的に作成する手法を提案する。具体的には動画とスライドの組から、スライド中の単語出現状況や提示時間、シーンの順序を使った複数の手法によるダイジェスト作成を行う。また、被験者により選択された正解セットと比較することで、それぞれの手法の結果を評価し、有効性を検証する。

以下には本稿の構成を述べる。2.節において関連研究について議論し、3.節でMPMeisterを説明する。4.節で提案手法の説明を、5.節で評価実験を行う。最後に6.節においてまとめと今後の課題を述べる。

2. 関連研究

2.1 テキスト要約

テキスト要約は主に二つの手法がある。一つ目は利用者の問い合わせによる要約手法である。Sandersonの研究[3]では、ある文書から利用者の問い合わせに最も適合する文を抽出す

る。二つ目は文書内容のトピックをよく表す文を抽出する手法である。Yihongらの研究[4]では、基の文書に対し、単語の出現頻度の考慮と潜在的意味解析を行い、重要文を抽出することで、自動要約を作成する手法を提案している。また、最近、Barzilayらはテキスト内容での語彙の連鎖を用い、自動要約する研究[5]を行っている。

本研究では、講義講演コンテンツを対象とする為、このようなテキスト要約の成果はそのまま適用できない。しかし、テキスト要約で用いられている技術を応用することができる。

2.2 音声要約

音声自動要約の分野には様々な研究が行われている。その中で堀の研究[6]では、音声認識を用い、単語抽出により発話文における重要な情報をできる限り保持し、日本語としては意味が理解できる要約文を生成する手法を提案している。

本研究では、まだ講義講演コンテンツの音声情報を利用していないが、今後の課題として、音声情報を利用することが必要であると考えている。

2.3 動画要約

最近では、スポーツ動画に対してダイジェストを作成する研究も行われている。Jinjunら[7]はスポーツ動画と音声の特徴を発見し、動画・音声のキーワードを作成し、同期化する。次に、これらのキーワードの組み合わせベクトルから、隠れマルコフモデル分類器で、重要なイベントを抽出するフレームワークを提案している。Marcus J.P.らのANSES [8]はニュース動画要約のため、動画ヒストグラム解析を行い、キーフレームを発見すると共に、字幕から取られるテキスト情報を用い、重要部分を抽出する。しかし、講義講演の動画を対象とした場合は、スポーツ動画の得点シーンのような特徴的なシーンに当たらない、あるいはニュース動画要約のようにアンコモンシーンがないため、これらの手法は適用できないと考える。

本研究と非常に近いのはLiweiらの研究[9]である。彼らもスライドを使用した音声・動画プレゼンテーションを自動的に要約することを目的としている。この研究ではシーンの提示時間、新しい話題の変化を発見するためのプレゼンターの声の高さの解析と、利用者のビデオの各時刻に対するアクセス頻度を基に、重要な箇所を抽出し、ダイジェストを作成する。但し、彼らはスライド中のテキスト内容を考慮しないので、その点で本研究と異なる。

これまで、我々は講義講演コンテンツから、検索キーワードに合致するシーンを探すために、スライド中の単語情報だけでなく、スライド提示の時間情報や順序情報を用いたUPRISE [2]を提案してきた。また[10]において手法を実装し、その後[11]では講演者の音声情報を考慮した改良と、[12]ではレーザーポインタ情報等を考慮した改良を行った。しかし、ダイジェストの自動作成では、講義講演のトピックを発見する必要がある。そのため、UPRISEで手法をそのまま本研究に適用することができないが、UPRISEで用いられる算出式は本研究でも有効であると期待できる。

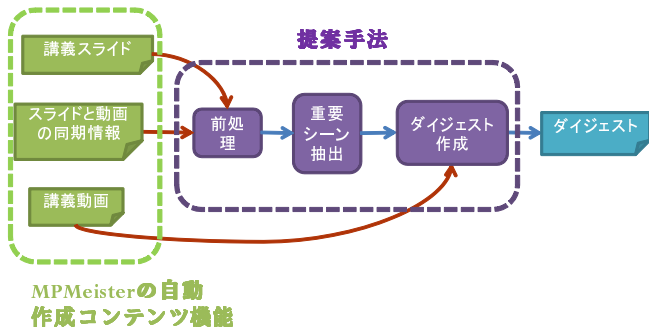


図2 ダイジェスト自動作成の概要

3. MPMeister

MPMeister は (株) リコーによって開発され、Microsoft の PowerPoint を用いた動画と PowerPoint ファイルから、それらを統合したウェブコンテンツを簡単に作成することができるソフトウェアである。MPMeister は PowerPoint スライドショーと共に起動することにより、PowerPoint でのスライドを切り替えるタイミング取得し、動画とスライドを同期させたコンテンツを生成する。作成したコンテンツを使用する際は、シーン毎の目次、動画とスライドをウェブ画面に提示し、ユーザが指定したシーンの動画とスライドに画面を切り替える機能を提供する。

MPMeister はコンテンツ生成と同時にコンテンツの内容を記述した MPEG-7 ファイル (XML 形式) を出力する。本研究では、MPEG-7 ファイルをメタ情報として利用する。その MPEG-7 ファイルには各シーンに出現するスライドのテキスト内容とシーンの提示時間が含まれている。

4. 提案手法

本稿では、図2のように、前節で説明した MPMeister の MPEG-7 ファイルに対し、講義スライドとスライドと動画の動機情報を用い、前処理を行う。次にシーンの重要度を算出し、重要シーンを抽出する。それを基に、ダイジェストを自動作成する手法を提案する。

4.1 前処理

まず、収録された講義からスライドのテキスト情報とスライドの提示時間を収集する。次に、講義講演でスライド移動やプレゼンターの操作ミスなどを考慮し、提示時間が3秒以下のシーンを削除する。その後、日本語形態素解析システム Sen [13] で、スライドのテキスト内容から意味のある単語を抽出する。最後に、J-E Ontology [14] を用い、意味の同じ英語と日本語の単語を発見し、同じ単語として扱うようにする。

4.2 重要シーン抽出

4.2.1 前提

システムは複数の講義の録画を保持するが、ダイジェスト作成の対象はその中の一つの講義 L_l とする。と同時に、システムはその講義で用いられた全スライド $S = \{s_1, s_2, \dots, s_N\}$ を持つ。ここでは、スライド中の単語出現にのみ着目することから、各スライド s_i は、単語の並び $s_i = [v_{i_1}, v_{i_2}, \dots, v_{i_m}]$ と

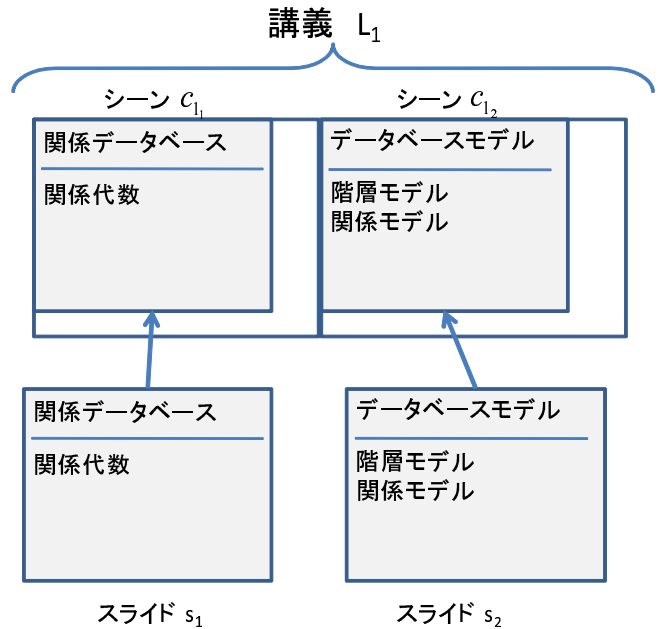


図3 講義 L_1 例

して表現される。ここで、全スライド S 中の全単語の集合を $W = \{w_1, w_2, \dots, w_M\}$ とすると、当然 $v_{i_j} \in W$ となる。この時、講義 L_l の録画はスライドの切り替えによるシーンの並びとして $L_l = [c_{l_1}, c_{l_2}, \dots, c_{l_n}]$ として表現する。ここで、シーン c_{l_k} は S 中のいずれかのスライド s_i に対応することから $c_{l_k} = [v_{i_1}, v_{i_2}, \dots, v_{i_m}]$ として扱うことができる。なお、実際の講義では、同じスライド s_i が再利用やバックトラック等で、異なるシーン中に複数回現れることに注意する必要がある。

簡単な例を図3に示す。ある講義 L_1 はスライド s_1 を用いるシーン c_{1_1} とスライド s_2 を用いるシーン c_{1_2} からなると仮定する。このとき、

$$\begin{aligned}
 L_1 &= [c_{1_1}, c_{1_2}], \\
 S &= \{s_1, s_2\}, \\
 s_1 &= [v_{1_1}, v_{1_2}, v_{1_3}, v_{1_4}], \\
 s_2 &= [v_{2_1}, v_{2_2}, v_{2_3}, v_{2_4}, v_{2_5}, v_{2_6}], \\
 v_{1_1} &= \text{“ 関係 ”}, v_{1_2} = \text{“ データベース ”}, v_{1_3} = \text{“ 関係 ”}, \\
 v_{1_4} &= \text{“ 代数 ”}, \\
 v_{2_1} &= \text{“ データベース ”}, v_{2_2} = \text{“ モデル ”}, v_{2_3} = \text{“ 階層 ”}, \\
 v_{2_4} &= \text{“ モデル ”}, v_{2_5} = \text{“ 関係 ”}, v_{2_6} = \text{“ モデル ”}, \\
 W &= \{ \text{“ 関係 ”}, \text{“ データベース ”}, \text{“ 代数 ”}, \text{“ モデル ”}, \text{“ 階層 ”} \} \\
 &\text{となる。}
 \end{aligned}$$

4.2.2 シーンの重要度算出

シーンの重要度を算出する五つ算出式を提案する。その際、「何度も繰り返し言及される概念は重要な概念である」という仮定 [15] の元に、シーン中の単語の出現状況に着目して、重要度を算出する。

以下に五つの算出式を説明するが、各算出式ではいづれもスライドの文章構造を考慮した単語への重み付け $p(v_{i_j})$ を用いる。 $p(v_{i_j})$ は以下の様に定義する。

$$p(v_{i_j}) = \begin{cases} \rho_t & v_{i_j} \text{ がタイトルに出現} \\ \rho_{b_l} & v_{i_j} \text{ がインデントレベル } b_l \text{ に出現} \\ 0 & v_{i_j} \text{ が出現しない} \end{cases} \quad (1)$$

a) シーンの提示時間を考慮した算出式 I_d

スライドの提示時間が長ければ、その中の概念について詳しく説明していて、シーンの重要度が高いと推測できる。そこで、UPRISE [12] と同様に、シーンの提示時間を考慮した算出式 I_d を提案する。

$$I_d(c_{l_k}, \theta) = I_p(c_{l_k}) \cdot t(c_{l_k})^\theta \quad (2)$$

ここで、 $t(c_{l_k})$ はシーン c_{l_k} の提示時間を表わし、 θ は時間の影響を決めるパラメータである。また、上で用いられる I_p は以下のように求められる。

$$I_p(c_{l_k}) = \sum_{w_x \in c_{l_k}} \left(\sum_{c_y \in L_l} \left(\sum_{v_z \in c_y, v_z = w_x} p(v_z) \right) \right) \quad (3)$$

例として、上の図 3 に示された講義 L_1 の二つのシーンの重要度を算出する。現在の MPMeister の制約からスライドの本文のインデントレベルが一つしかないとし、 $\rho_t = 5$ 、 $\rho_{b_l} = 1$ とする。すると、 c_{1_1} において“関係”(v_{1_1} に当たる) という単語はタイトルに含まれるため $p(v_{1_1}) = 5$ となる。また“代数”(v_{1_4} に当たる) という単語は本文に含まれるため、 $p(v_{1_4}) = 1$ となる。以下同様に計算すると $I_p(c_{1_1}) = 18$ 、 $I_p(c_{1_2}) = 25$ となる。

b) シーンの順序を考慮した重要度 I_{dc}

あるシーンに出現する単語が前後のシーンにも出現する場合、そのシーンは単語を説明していると思われるため、重要であると考えられる。そこで、UPRISE [12] と同様の算出式 I_{dc} を用いる。

$$I_{dc}(c_{l_k}, \theta, \delta, \varepsilon_1, \varepsilon_2) = \sum_{j=\gamma-\delta}^{j=\gamma+\delta} I_d(c_{l_k}, \theta) \cdot E(j - \gamma, \varepsilon_1, \varepsilon_2) \quad (4)$$

δ は前後何枚まで影響を与えるかを定めるパラメータで、 $E(j - \gamma, \varepsilon_1, \varepsilon_2)$ は前後関係の影響を決める関数で、以下のように定義される。

$$E(x, \varepsilon_1, \varepsilon_2) = \begin{cases} \exp(\varepsilon_1 x) & (x < 0) \\ \exp(-\varepsilon_2 x) & (x \geq 0) \end{cases}$$

この重要度によって、重要度はそのシーンの前後 θ だけの範囲の影響を受け、 ε が小さいほど影響を受けやすくなる。例えば、 $\theta = 4$ 、 $\varepsilon_1 = 5.0$ 、 $\varepsilon_2 = 0.5$ の時、そのシーンの重要度が前後 4 シーンの重要度に影響を受け、 $\varepsilon_1 > \varepsilon_2$ であるため後に続くシーンの方により強い影響を受ける。

c) 単語の特定性を考慮した重要度 I_{dr}

単語自体の重要性を考える場合、idf 等と同様に、他のシーンでその単語がどの程度出現しているかを考慮することも有効であると思われる。このため、以下のように単語の特定性を考

表 1 各算出式の考慮した点のまとめ

シーン重要度の算出式	考慮した点				
	単語の出現頻度	シーンの提示時間	シーンの提示順序	単語の特定性	スライド中の出現単語数
I_d					
I_{dc}					
I_{dr}					
I_{drc}					
I_{df}					

慮した算出式 I_{dr} を提案する。

$$I_{dr}(c_{l_k}, \theta) = \sum_{w_x \in W} \frac{app(c_{l_k}, w_x)}{\sum_{c_y \in L_l} app(c_y, w_x)} \cdot I_{pl}(c_{l_k}) \cdot t(c_{l_k})^\theta \quad (5)$$

なお上での算出式にある I_{pl} は以下のように求める。

$$I_{pl}(c_{l_k}) = \sum_{v_x \in c_{l_k}} p(v_x) \quad (6)$$

また、 $app(c_{l_k}, w_x)$ はシーン c_{l_k} 中の単語 w_x が何回出現するかを計算する関数である。

d) 特定性とシーンの順序を考慮した重要度 I_{drc}

上記の組み合わせとして特定性とシーンの順序を考慮した算出式 I_{drc} を提案する。

$$I_{drc}(c_{l_k}, \theta, \delta, \varepsilon_1, \varepsilon_2) = \sum_{j=\gamma-\delta}^{j=\gamma+\delta} I_{dr}(c_{l_k}, \theta) \cdot E(j - \gamma, \varepsilon_1, \varepsilon_2) \quad (7)$$

e) スライド中の出現単語の数を考慮した重要度 I_{df}

シーン中の単語数が多ければ、上で提案した算出式による値は大きくなるが、出現単語の数が少ない方がそのシーンの重要度が高い可能性がある。そこで、各単語の出現頻度を考慮した算出式 I_{df} を提案する。

$$I_{df}(c_{l_k}, \theta) = \frac{1}{\sum_{w_x \in W} app(c_{l_k}, w_x)} \cdot I_d(c_{l_k}, \theta) \quad (8)$$

各算出式の意味を表 1 にまとめた。

4.2.3 重要シーン抽出方法

各算出式を適用し、シーンの重要度を算出する。平均値以上のシーンを講義講演コンテンツの重要シーンとする。

4.3 ダイジェスト自動作成

まず、各重要シーンの重要度に比例した時間を、ダイジェストに入れる時間と定める。次に、各重要シーンの先頭からダイジェストに入れる時間のみを切り取ったものを繋げてダイジェストを作成する。これは、各重要シーンの途中の抽出ではシーン内トピックの切れ目と合致せず不自然なダイジェストが作成されてしまうためである。我々の予備実験では、各重要シーンの先頭三分の一時点から作成したダイジェストはつながりが不

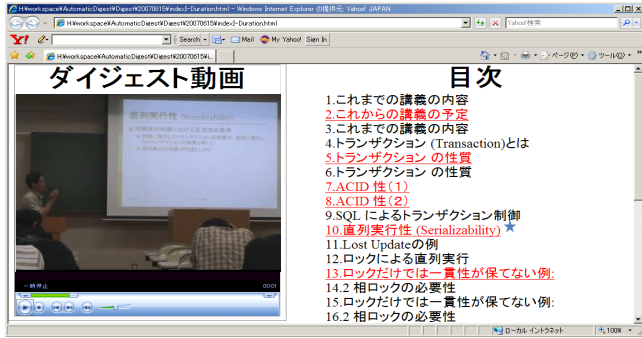


図 4 ダイジェスト再生の画面

表 2 各講義の特徴

講義	講義内容	シーン数	単語数	平均単語数	平均提示時間
講義 A	関係度数	59	1548	26	85 秒
講義 B	トランザクション	32	1043	33	165 秒
講義 C	イントロダクション	31	1674	54	160 秒
講義 D	データベースの内部構造	44	1286	29	110 秒
講義 E	問い合わせ処理と前授業の復習	98	4256	43	48 秒
講義 F	オブジェクト指向データベースと前授業の復習	67	3306	49	68 秒

自然であり、被験者評価が著しく低下することを確認した。なお、重要シーン中の自然なサブシーン抽出とダイジェストへの利用は今後の課題とする。最後に、ユーザがダイジェストを見られるように、各重要シーンの開始時間、提示時間とビデオファイルへのリンクというメタデータを Windows Media メタファイル形式のファイルを作成する。また、自動ダイジェストを再生する時の画面を図 4 に示す。右側に表示する目次で、下線が付いている項目が重要シーンとして抽出されるものである。

5. 評価実験

本研究の目的は講義講演ビデオから、自動ダイジェストを作成することであるが、今回は講義講演コンテンツから重要シーンを抽出する所までの評価を二つの被験者実験により行った。一つ目は提案された五つの算出式の有効性を検証する被験者実験であり、二つ目はこれらの算出式に用いる二つのパラメータであるタイトルに出現する単語の重みの ρ_t とシーンの提示時間の θ を変化させ、これらのパラメータの影響の度合いを評価する被験者実験である。

5.1 実験 1：提案手法の有効性の検証

5.1.1 実験環境

実験では実際に大学で行われた二つの講義 A と講義 B を用いる。これらの講義の内容、シーン数、単語数などの特徴を表 2 に示す。

5.1.2 実験方法

二つの講義を見たことがない 6 名の被験者にダイジェストに含めたい場面を含む重要シーンを選んでもらう。次に 6 名の内 4 名が重要シーンとしたシーンを正解シーンとし、正解シーン

表 3 各講義に対する各算出式の再現率、適合率と F 尺度

算出式	講義 A			講義 B		
	再現率	適合率	F 尺度	再現率	適合率	F 尺度
I_d	0.45	0.31	0.37	0.67	0.40	0.50
I_{dc}	0.36	0.24	0.29	0.83	0.42	0.56
I_{dr}	0.45	0.33	0.38	0.50	0.33	0.40
I_{drc}	0.45	0.22	0.30	0.50	0.25	0.33
I_{df}	0.45	0.22	0.30	0.83	0.42	0.56

の集合を C とする。そして C と各算出式により抽出された重要シーンから F 尺度を算出した。

F 尺度は情報検索システムの性能評価でよく使われている再現率と適合率の調和平均である。F 尺度が高ければ、性能が良いことを意味する。それぞれ、以下の様に定義される。

$$\text{再現率} = \frac{|\text{算出式による重要シーン集合} \cap \text{正解シーン集合}|}{|\text{正解シーン集合}|}$$

$$\text{適合率} = \frac{|\text{算出式による重要シーン集合} \cap \text{正解シーン集合}|}{|\text{算出式による重要シーン集合}|}$$

$$F \text{ 尺度} = \frac{2 \cdot \text{再現率} \cdot \text{適合率}}{\text{再現率} + \text{適合率}}$$

また各式 (2), (4), (5), (7), (8) のパラメータは $\rho_t = 5, \forall b_l \rho_{b_l} = 1 \theta = 1 \delta = 3 \epsilon_1 = 5 \epsilon_2 = 0.5$ を用いた。

5.1.3 実験結果と考察

各講義での各算出式の再現率、適合率および F 尺度を表 3 に示す。被験者による重要シーンと各算出式による重要シーンの食い違いを図 5 と図 6 に示す。

図 5 と図 6 の被験者による重要シーンにおいて、各被験者によって選ばれたシーンに多くの共通部分が見られることから、利用者毎に講義のダイジェストとして見たいシーンが異なる事は少なく、ダイジェストに含むべきシーンが存在することが確認できた。

図 5 と図 6 において、正解シーンがある程度提案手動により見つけれられたため、本稿で提案した重要シーンの抽出方法はダイジェスト作成に当たり、有効であると考えられる。

表 3 において、講義 A と講義 B で、F 尺度で最も高い値を示している算出式が異なることから、今回の実験ではどの算出式が最も良いか判断できなかった。そのため、他の講義を用い、評価実験を行う必要がある。

図 5 の講義 A では被験者が重要シーンと選んでいるにもかかわらず、提案手法では重要シーンとできなかったシーンが幾つかある。理由を調べてみたところ、重要シーンとできなかったシーンの多くは図を含んでいるものが多かった。MPMeister では図に含まれている単語は抽出できないため、重要度が下がり重要シーンに選ばなかったと考えられる。よって、今後は図中からも単語を抽出する手法を考案する必要がある。

図 6 の講義 B のシーン 32 は、提案手法では重要シーンとしているが、被験者は重要シーンとしていない。この理由を調べたところ、シーン 32 はスライドに表示した問題を、11 分間の間生徒に解かせるシーンであった。全ての重要度算出方法にお

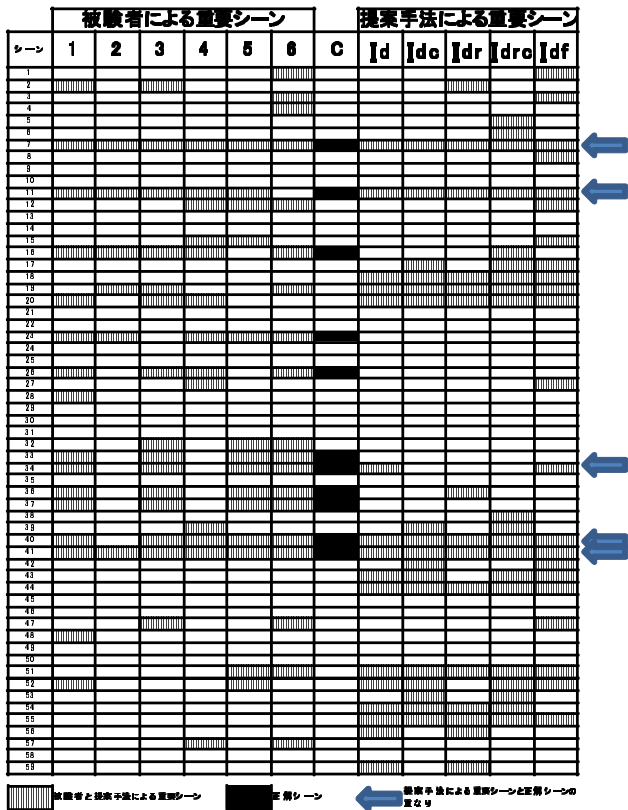


図 5 被験者と提案手法による重要シーンの食い違い - 講義 A

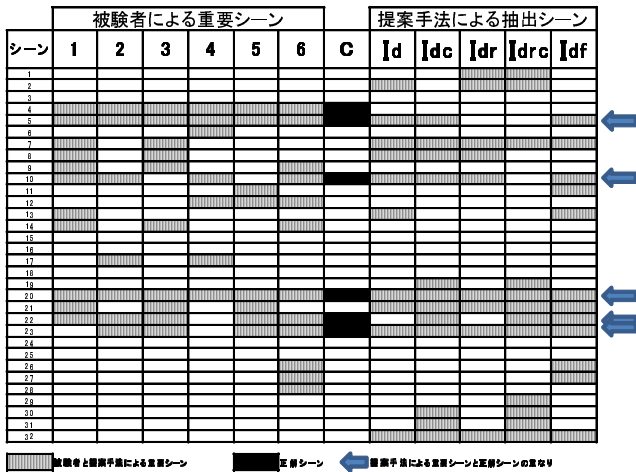


図 6 被験者と提案手法による重要シーンの食い違い - 講義 B

いて、スライドの提示時間を考慮しているため、11 分間という長時間提示されていたこのシーンが、重要シーンと計算されたと思われる。そのため、このような長時間授業が進まないシーンを除去することが、手法の有効性を向上すると考えられる。このことを確認するために追加実験を行った。

「練習問題」シーンを対象から外した追加実験の結果は表 4 と図 7 のようになった。表 4 を見ると値が変わらない I_{df} 以外にすべての F 尺度が向上している。また、図 7 を見ると、図 6 の講義 B よりも被験者による重要シーンを多く含めている。この結果から、長時間授業が進まないシーンの除去により、手法

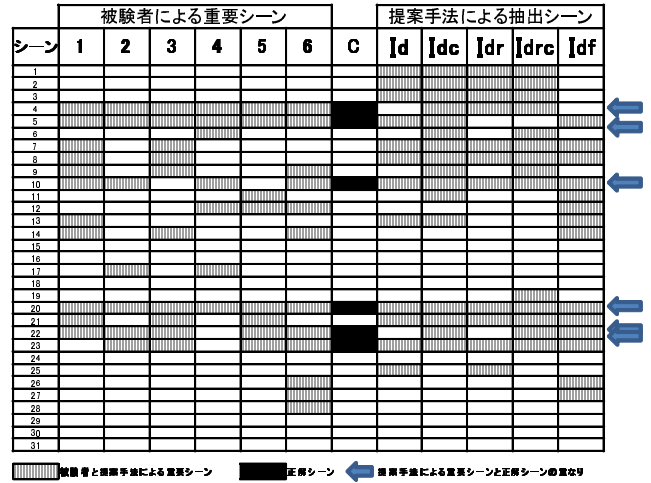


図 7 被験者と提案手法による重要シーンの食い違い - 講義 B (「練習問題」シーン削除後)

表 4 講義 B に対する各算出式の再現率、適合率と F 尺度 (「練習問題」シーン削除前と後)

算出式	削除前			削除後		
	再現率	適合率	F 尺度	再現率	適合率	F 尺度
I_d	0.67	0.40	0.50	0.83	0.45	0.59
I_{dc}	0.83	0.42	0.56	1.00	0.43	0.60
I_{dr}	0.50	0.33	0.40	0.67	0.44	0.53
I_{drc}	0.50	0.25	0.33	0.83	0.38	0.53
I_{df}	0.83	0.42	0.56	0.83	0.42	0.56

の有効性が向上することを確認できた。よって、今後はこのようなシーンを除去する手法を考案していく必要がある。

5.2 実験 2 : 考慮した点の影響の度合いの評価

式 (1) のタイトルに出現する単語の重みを表すパラメータ ρ_t とシーンの提示時間を表すパラメータ θ を変更させ、提案された 5 種類の算出式への影響を評価するため、被験者実験を行った。

5.2.1 実験環境

実験 1 で使用された二つの講義と異なる実際に行われた四つの講義 C、講義 D、講義 E、講義 F を利用し、被験者実験を行った。この四つの講義の内容、シーン数、単語数等の特徴を表 2 に示す。

5.2.2 実験方法

実験 1 と同じように、講義ビデオをまだ見ていない 2 名の被験者に講義ビデオからダイジェストに含めたい場面を含む重要シーンを選んでもらった。2 名共に重要シーンとしたシーンを正解シーンとして、正解シーンの集合を正解セットとした。

次に正解セットから、パラメータ ρ_t を 1 から 10 まで 1 刻みで変化させ、講義毎で各算出式による再現率、適合率と F 尺度を算出した。最後に、各講義で F 尺度を最も良い結果とさせた ρ_t を用い、 θ を 0 から 5 まで 0.5 刻みで変化させ、各算出式による再現率、適合率と F 尺度を算出した。その他のパラメータを変更しなかった。

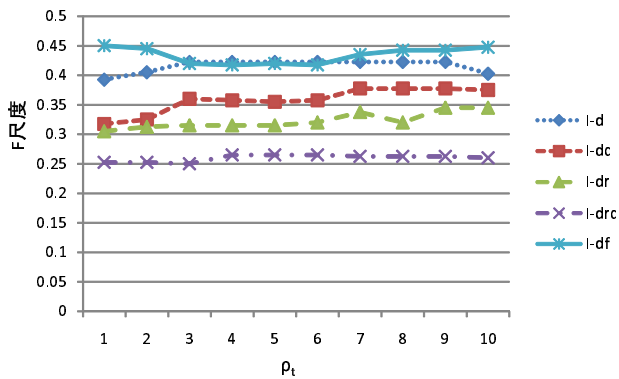


図 8 ρ_t の変化と F 尺度の平均値の推移

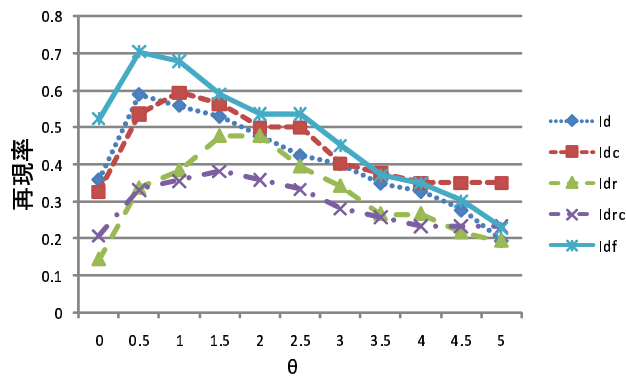


図 10 θ の変化と再現率の平均値の推移

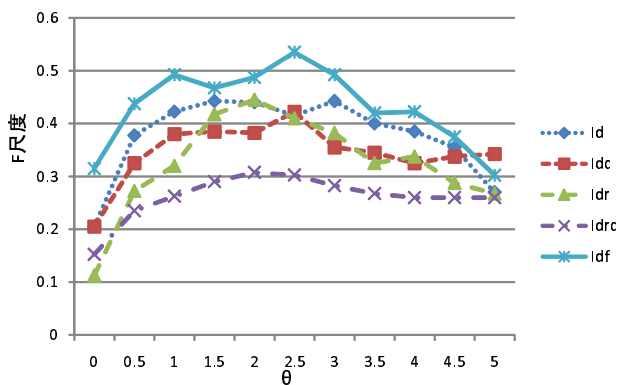


図 9 θ の変化と F 尺度の平均値の推移

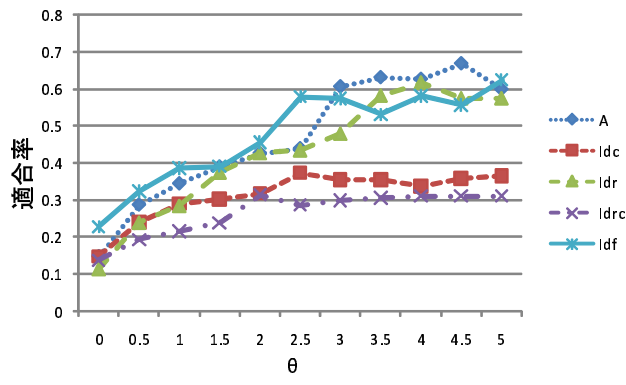


図 11 θ の変化と適合率の平均値の推移

5.2.3 実験結果と考察

ρ_t と θ の値に対応する四講義の F 尺度の平均値の推移を図 8 と図 9 に示す。

図 8 のグラフから分かるように、多くの算出式で ρ_t が大きいほど F 尺度が大きいため、タイトルに出現する単語をより重要とすることが有効であると考えられる。しかし全体としては、結果に影響が小さいことが分かった。

次に、図 9 に表示されているグラフから分かるように、全部の算出式に対し、シーンの提示時間を考慮した F 尺度が考慮しない方 $\theta = 0$ と比べ良い結果が出た。それによって、シーンの提示時間を考慮することが有効であると判明した。

ところが、 θ を大きくすれば、全ての算出式による F 尺度が低くなる傾向が見られた。それで、シーンの提示時間を考慮するべきであるが、パラメータを大きくしすぎない方がよいと考える。

また、 θ が提案される算出式に対する再現率と適合率に及ぼす影響を考察した。 θ の変化に対する四つの講義の再現率と適合率の平均値の推移を図 10 と図 11 に示す。

この結果から再現率に関して、全ての算出式は $\theta > 0$ の時、 θ が大きいほど、良くなった。一方、適合率に関しては、 θ が大きいほど、良くなった。理由としては、 θ が大きくなると、各シーン重要度には差が大きくなるため、平均値以上の重要度を持つシーンの数が少なくなると考える。

最後に、図 8 と図 9 にスライド中の出現単語の数を考慮した

算出式 I_{df} による F 尺度が良かったから、重要シーンを抽出することに関して、スライド中の出現単語を考慮することが重要であると考えられる。

6. まとめと今後の課題

講義講演コンテンツのダイジェスト自動作成に既存の動画要約研究をそのまま適用することはできない。本稿では講義講演のビデオのダイジェスト自動作成の第一歩として、スライド中の単語の出現状況、シーンの提示時間、順序を用い各シーンの重要度を算出する複数の手法を提案し、それを基に、ダイジェストを自動作成する手法を提案した。また、重要シーン抽出手法の評価を実際の講義を用いた被験者実験により行った。その結果、正解シーン集合と抽出した重要シーン集合に多くの共通部分が見られたため、提案手法が有効であると考えられる。また、重要シーン抽出のため、単語の出現分布とシーンの提示時間が重要であることが判断できた。

今後の課題として以下の四つが挙げられる。

第一にシーン重要度のさらなる評価である。今回、シーンの重要度を算出するための 5 種類の算出式の性能をより正確に評価するため、他の授業の講義で実験していきたいと考える。

第二に提案手法の改良である。今回、単語の出現状況のみを考慮したが、今後、日本語のオントロジーの適用について考える必要がある。また、今回の実装システムは前処理の段階では、スライドのテキスト内容から意味のない単語、及び提示時間が 3 秒以下のシーンを削除しているが、音声情報追加などで各講

義において本質に対応するフィルタリング手法を考案する必要がある。

第三に重要シーンから PowerPoint のアニメーション機能による情報や音声情報等を用い、本来ダイジェストに含まれるべきサブシーンを抽出する必要がある。今回は重要シーンの最初の部分を切ったものを繋げるによりダイジェストを生成したが、実際にダイジェストに入れるべき部分は必ずしも最初の部分とは限らないためである。

第四にダイジェスト再生のインタフェースの改良である。現在は画面にはビデオとダイジェストの目次しか表示されていないが、将来はダイジェストの長さやビデオに出現されるスライドの内容の表示等、多様な利用者の要求に対応できるようなインタフェースを作成する必要がある。

謝 辞

なお、本研究の一部は、文部科学省科学研究費補助金特定領域研究(19024028)、独立行政法人科学技術振興機構 CREST、および東京工業大学 21 世紀 COE プログラム「大規模知識資源の体系化と活用基盤構築」の助成により行われた。

文 献

- [1] Ricoh Japan, “MPMeister II — Ricoh Japan,” <http://www.ricoh.co.jp/mpmeister>.
- [2] H. Yokota, T. Kobayashi, T. Muraki, and S. Naoi, “UPRISE: Unified Presentation Slide Retrieval by Impression Search Engine,” *IEICE Trans. on Info. and Syst.*, vol.E87-D, no.2, pp.397-406, 2 2004.
- [3] M.Sanderson, “Accurate User Directed Summarization from Existing Tools,” *Proc. of the 7th International Conference on Information and Knowledge Management (CIKM98)*, 1998.
- [4] Y. Gong, and X. Liu, “Generic Text Summatization Using Relevance Measure and Latent Semantic Analysis,” *Proc. of the 24th annual international ACM SIGIR conference on Research and development in information retrieval*, pp.19-25, 2001.
- [5] R. Barzilay, and M. Elhadad, “Using Lexical Chains for Text Summarization,” *Proc. of the Intelligent Scalable Text Summarization Workshop (ISTS'97)*, 1997.
- [6] 堀智織, A Study on Statistical Methods for Automatic Speech Summarization, Ph.D thesis, Tokyo Institute of Technology, 2002.
- [7] J. Wang, C. Xu, E. Chng, and Q. Tian, “Sport Highlight Dectection from Keyword Sequences using HMM,” *Proc. of IEEE International Conference on Multimedia and Expo, ICME '04*, vol.1, pp.599-602, 2004.
- [8] M.J. Pickering, L. Wong, and S.M. Rüter, *Image and Video Retrieval*, Springer Berlin / Heidelberg, 2003.
- [9] L. He, E. Snocki, A. Gupta, and J. Grudin, “Auto-Summarization of Audio-Video Presentation,” *Proc. of the seventh ACM international conference on Multimedia (Part 1)*, pp.489-498, 1999.
- [10] 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*, vol.J88-D-1, no.3, pp.715-726, 3 2005.
- [11] 仲野亘, 小林隆志, 直井聡, 横田治夫, 岩野公司, 古井貞熙, “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 回電子情報通信学会データ工学ワークショップ (DEWS2006), 2006.
- [12] 仲野亘, 小林隆志, 直井聡, 横田治夫, 古井貞熙, “講義講演シーン検索手法におけるレーザーポインタ情報と音声情報の粒

度を考慮した統合,” 第 18 回電子情報通信学会データ工学ワークショップ (DEWS2007) 論文集, pp.E1-3, 2007.

- [13] “形態素解析システム Sen,” <http://www.mlab.im.dendai.ac.jp/~yamada/ir/MorphologicalAnalyzer/Sen.html>.
- [14] “An Experimental Japanese - English Ontology,” <http://www.kanzaki.com/docs/sw/jwebont.html>.
- [15] H. Luhn, “A Statistical Approach to Mechanized Encoding and Searching of Literary Information,” *IBM Journal of Research and Development*, vol.1, no.4, pp.309-317, 1957.