講演動画アーカイブスからキーワードに合致する講演ダイジェストの自動作成

呉　　怡† 渡辺　陽介†† 横田　治夫††,†††

† 東京工業大学　開発システム工学科　〒 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　〒 152-8552 東京都目黒区大岡山 2-12-1
††† 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
E-mail: †{goi,watanabe}@de.cs.titech.ac.jp, ††yokota@cs.titech.ac.jp

あらまし　近年，学会などにおける講演が動画コンテンツとして多く蓄積されている．しかし，大量の講演コンテンツを全て閲覧することは多くの時間を要する．そこで，本研究では，こうした講演動画アーカイブスに対し，ユーザが短時間で興味のあるトピックスに関連する講演の概況を把握できるように，キーワードに合致する複数の講演コンテンツを含むダイジェストの提供を目指す．提案手法では，まず単語の出現頻度やその他の情報を用いて，講演コンテンツのキーワードに対する適合度を計算することにより，ダイジェストに含めるコンテンツを定める．更に，それらの講演コンテンツで使われるスライド構成や提示時間情報等を考慮しながら，ダイジェストに含めるシーンを抽出し，講演動画ダイジェストを自動作成する．

キーワード　E-learning，映像要約

Keyword-based Automatic Digest Generation from Multiple Presentation Video Archives

Yi WU†, Yousuke WATANABE††, and Haruo YOKOTA††,†††

† Department of International Development Engineering, Tokyo Institute of Technology
2–12–1 Ookayama, Meguro-ku, Tokyo, 152–8550 JAPAN
†† Global Scientific Information and Computing Center, Tokyo Institute of Technology
2–12–1 Ookayama, Meguro-ku, Tokyo, 152–8550 JAPAN
††† Department of Computer Science, Graduate School of Information Science and Engineering
2–12–1 Ookayama, Meguro-ku, Tokyo, 152–8550 JAPAN
E-mail: †{goi,watanabe}@de.cs.titech.ac.jp, ††yokota@cs.titech.ac.jp

Abstract　In recent years, a lot of presentation video contents has been accumulated. However, they often require very long time to view all presentations videos. In this study, we aim to provide a summary of the presentations whose topics interested to a user, and to generate a digest automatically including multiple presentation videos matching with the keywords. We consider several factors of presentations such as term frequency to calculate the relevance of the presentation content relating to the given keywords and determine the duration of the digest. Moreover, we consider structure of presentation slides, duration information for explaining slide to extract important scenes to be included in the digest.

Key words　E-learning, Video Digest

1. はじめに

講演・講義を行う際に，プレゼンテーションソフトウェアを用いて説明することが多く，近年では，学会における講演を活用するため，その様子を記録した講演動画アーカイブが大量に蓄積されている．例えば，日本データベース学会 [1] では過去に行なわれた研究シンポジウムにおける研究発表の様子を収録した動画アーカイブスを配信している．しかし，こうした講演動画アーカイブスに対する検索・要約のニーズが高まる中，それらを対象とした研究はまだ少ない．

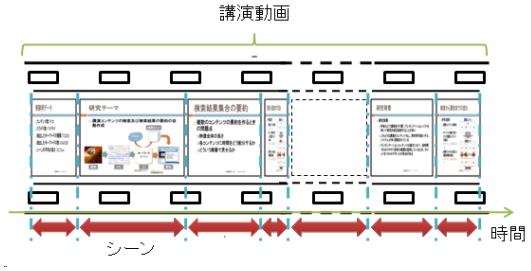

図 1　UPRISE におけるシーンの定義

このような背景から，これまで我々は主に講義コンテンツに注目し，講義で使われるスライド資料とそれの提示時間等を利用し，講義動画ストリームからキーワードに合致する部分を検索するシステムとして，UPRISE(Unified Presentation Slides Retrieval by Impression Search Engine) [2] を提案してきた．UPRISE では，スライドの切り替えタイミングを利用し，講義のビデオ映像をシーンに分割している (図 1)．そして，UPRISE におけるシーンの定義に基づき，ユーザが短時間で，講義コンテンツの内容を把握できるように，１つの講義コンテンツから，講義のトピックをよく表すシーンを抽出する手法 [3] を提案した．

しかし，膨大な動画データの中から検索キーワードに合致するシーンが検索できても，その検索結果を全て閲覧するには多くの時間を要する．また，キーワードを含むシーンだけ参照しては，講演における発表内容を十分に理解できない可能性が大きい．そして，講演では講義に比べ，時間が短く，個別の講演コンテンツよりも，複数の講演コンテンツをまとめて要約するニーズの方が多いと考えられる．

そこで本稿では，検索キーワードに基づく複数の講演コンテンツを対象としたダイジェストを自動作成する手法を提案する．まず，従来の重み付け手法を利用し，キーワードに合致する講演コンテンツを検索し，ランク付けをする．そして，ランク付けされたコンテンツから，重要なシーンを抽出するために講演に含む単語に重みをつけ，それの重みを利用し，シーンの重要度を算出する．重要度を算出する際には，単語の出現状況やスライドの提示時間，シーンの前後関係，そして講演のスライド資料に含む特徴的題目を考慮する．最後にランキングの結果を用いて，与えられたダイジェストの時間に納まるように時間配分を行い，ダイジェストを自動生成する．なお，講演の場合では，キーワードを含まないシーンでもそのキーワードに関連性を持つ可能性が大きいため，ダイジェスト用シーンを抽出するにあたって，講演資料に出現する全単語の出現状況を考慮している．

以下に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. 節では提案手法の説明を行う．4. 節でダイジェスト用シーンの抽出手法に関する評価実験を行い，最後に 5. 節においてまとめと今後の課題について述べる．

2.　関 連 研 究

2.1　ニュース・スポーツ動画の要約

これまで，動画要約に関する研究の中に，ニュースとスポーツ動画に着目したものが多くあった．

Yang らが提案した VideoQA システム [4] では，ニュース動画に対して，画像の特徴分析，音声認識，文字認識，およびインターネット上にある関連新聞記事を利用することで，ユーザの質問に対し，ニュース動画アーカイブスの中から，回答となるニュースの要約映像を生成する．

Bezerra らの研究 [5] では，サッカーの動画映像を分析し，色などの特徴から，画像変化のリズムを捉え，ショットを検出・分類することで，試合のサマリを自動作成している．

しかし，講義・講演動画では，ニュースのようなクローズキャプションがなく，またその発話には高い自発性を持っている．その上，映像上の変化が少ないため，ニュース・スポーツ動画で用いられている要約手法をそのまま講義・講演動画に適用できないと考えられる．

2.2　講義・講演の扱い

2.2.1　講義・講演コンテンツの検索

これまで我々は講義・講演コンテンツに対し，高度な検索機能を提供するシステム UPRISE [2] を提案し, [6] において実装を行い，その有効性を示した．

UPRISE では，主に講義で使用されるプレゼンテーション資料とその動画を統合することで，スライドの切り替えタイミングから「シーン」を定義している（図 1）．このように定義されたシーンは必ずある 1 枚のスライドに対応しているため，スライドにおけるテキスト情報から，動画シーンの検索を可能にした．また，UPRISE では，検索を行なう際に，単語の出現頻度のみならず，単語の出現位置，スライドの提示時間及びシーンの前後関係といったプレゼンテーションの特徴を利用している．

しかし，講演コンテンツでは講義にない多くの特徴を持っているため，それらを考慮した新たなアプローチが必要である．

2.2.2　１つの講義・講演の要約

レーらの研究 [3] では，講義・講演で使われるスライドにおける単語の出現頻度，特定性，提示時間，シーンの順序を利用し，１つの講義・講演動画から重要シーンを抽出する手法を提案している．しかし，検索機能を提供していない点は本研究と異なる．

また，藤井らの研究 [7] では，講演に対応する予稿論文の文書構造と表層情報を利用して，利用者が指定した論文の一部に対応する講演音声の要約を提供する手法を提案している．また，堀らの研究 [8] においては，講演音声の発話文から重要な単語を抽出し，接合することで，１つの講演に対する音声の自動要約を試みた．

本研究では，検索結果から複数の講演動画に対してダイジェストの自動作成手法を提案している．またキーワードに関連する講演の検索及びダイジェストを作成する際，講演に対応する論文及び音声認識の情報を考慮していないので，今後の課題としてそれらを考慮に入れたいと考えている．

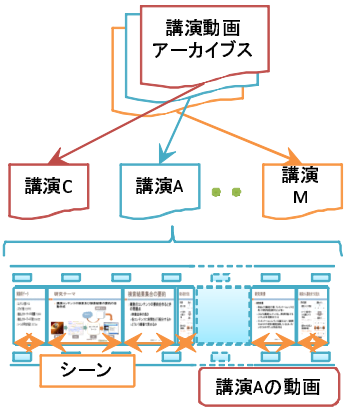

図 2　講演動画アーカイブの構成

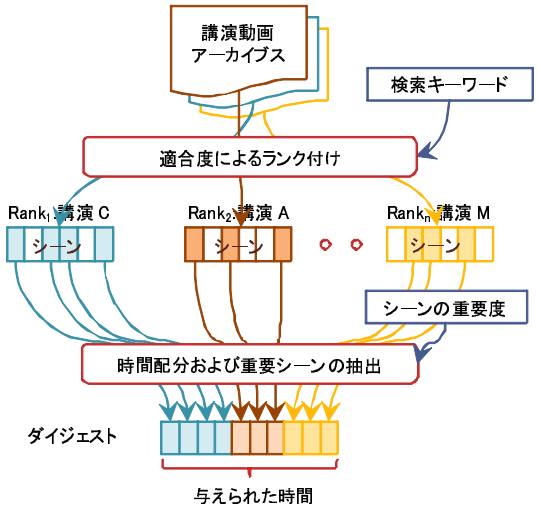

図 3　処理の流れ

3.　キーワードに基づく複数の講演を含むダイジェストの自動作成

3.1　概　　要

本稿では，講演の様子を収録した動画とそのときに使用されたプレゼンテーション資料及び両者の同期情報を保持している講演動画アーカイブ（図 2）を対象とする．

処理の流れとして，まず全ての講演コンテンツに対し，プレゼンテーションスライド中に含まれた検索キーワードの出現頻度に基づき，キーワードに適合する講演コンテンツをランク付けする．

次に，シーンの重要度を算出し，ランキング上位にある講演コンテンツから重要なシーンを抽出して，与えられた時間に納まるダイジェストを作成する．

3.2　前　処　理

前処理として，バックトラックや操作ミスによって生じたノイズシーンの影響を排除するために，講演動画における 1000msec 以下のシーンを削除する．その後，講演で使用されたスライドのテキストに対し，形態素解析を行ない，意味のある単語を抽出して索引を作る．

3.3　記号の説明

以降では，講演動画アーカイブスの集合を $P = \{p_1, p_2, \cdots, p_T\}$ とする．また，ある講演 p で使用されたスライドの集合を $S = \{s_1, s_2, \cdots, s_L\}$ とし，スライドの切り替えによって定義されたシーン（図 1）の集合を $C = \{c_1, c_2, \cdots, c_N\}$ で表す．そして，単語の集合を $T = \{t_1, t_2, \cdots, t_M\}$ とする．

3.4　キーワードに基づく講演動画コンテンツの検索

本研究では，キーワードに合致する複数の講演コンテンツを含むダイジェストを作成するために，まず全講演コンテンツに対してキーワード検索を行ない，ダイジェストに含める講演の候補を決める．ここでは，テキスト検索によく用いられる重み付け手法 tf.idf [9] を使い，キーワード k に対する講演 p_i の適合度 $score_i$ をこのように定義する．

$$score_i(k) = tf(k, p_i) * ipf(k) \tag{1}$$

ただし，

$$tf(k, p_i) = \frac{app(k, p_i)}{\sum_{t_j \in p_i} app(t_j, p_i)} \tag{2}$$

$$ipf(k) = \log \frac{|P|}{pf(k)} + 1 \tag{3}$$

とする．ここで，$app(k, p_i)$ はキーワード k が講演 p_i における出現頻度 (Term Frequency) で，$|P|$ はシステムが保持している全講演コンテンツの数を表し，$pf(k)$ はキーワード k が出現する講演の頻度 (Presentation Frequency) である．

3.5　シーン重要度の算出

以下において，前述の手法で見つかったキーワードに適合する講演から，それぞれの内容を把握するのに必要なシーンを抽出する手法について説明する．提案手法では，スライドにおける単語の出現頻度やそのほかの情報を利用し，単語に対する重みの算出し，さらに単語の重みに加え，スライドの提示時間，シーンの前後関係および講演スライド資料の特徴的題目を考慮し，シーンの重要度を算出する．

3.6　講演 p における単語 t の重み付け手法

提案手法では tf.idf [9] を基に，講演で使用するスライドの特徴を考慮し，単語の重みを算出する．

（1）　スライドの階層構造を考慮した算出式

スライドは大きく分けてタイトルと本文の 2 つの部分からなっている．タイトルに出現する単語がより重要で，講演のトピックを表しているものであるという仮定から，単語 t の出現位置を考慮した重み w_p の算出式を以下のように定義する．

$$w_p = tf(t, p) * isf(t) \tag{4}$$

ここでは，$tf(t, p)$，$isf(t)$ を以下のように求める．

$$tf(t, p) = \sum_l Position(l) * Count(t, p, l) \tag{5}$$

$$isf(t) = \log \frac{|S|}{sf(t)} + 1 \tag{6}$$

ただし，$Count(t, p, l)$ は講演 p の位置 $l(l \in \{$ タイトル，本

文 }) における単語 t の出現であり，S は講演 p に使われた全スライド集合で，$|S|$ はその要素数を表している．$sf(t)$ はスライド集合 S の要素のうち，単語 t が出現するスライドの頻度 (Slide Frequency) である．また，$Position(l)$ を以下のように定義する．

$$Position(l) = \begin{cases} \rho & (\text{l} = \text{タイトル}), \\ 1 & (\text{l} = \text{本文}). \end{cases} \tag{7}$$

（2） 講演で頻繁に言及される単語を重視する算出式

講演中のより多くのスライドに出現し，かつその講演について特定性のある単語は講演のトピックを説明する上で重要な概念であると考え，単語 t の重み w_s を以下のように定義する．

$$w_s = sf(t,p) * ipf(t) \tag{8}$$

ただし，$sf(t,p)$ は単語 t が講演 p におけるスライド頻度で，$ipf(t)$ の定義は式（3）に従う．

3.7 シーンの重要度算出

ここでは前述の算出式のいずれかを用い，講演 p における単語 $t(t \in p)$ の重み $w(t)$ を計算することを前提とする．

- 単語の重みのみ考慮したスライドの重要度 I_t

重要な単語をより多く含むスライドが重要であるという仮定のもとで，文献 [3] と同様にスライドの重要度はそこにある全単語の和として表し，I_t をこのように求める．

$$I_t(c_j) = I_t(s_i) = \sum_{t \in s_i} w(t) \tag{9}$$

ここでは，シーン c_j はスライド s_i に対応する．

- スライドの提示時間を考慮した重要度 I_d

より長い時間をかけて説明するスライドが重要であると考えられる．一方，説明時間が短ければ，単位時間に得られる情報量が大きいという考え方もできる．そこで，スライドの提示時間を考慮し，シーン c_j に対応するスライド s_i の提示時間が $T(c_j)$ 分間としたときに，I_d をこのように求める．

$$I_d(c_j, \theta) = (T(c_j) + 1)^{\theta} * I_t(s_i) \tag{10}$$

ただし，θ は提示時間の影響度を決めるパラメータである．

- シーンの前後関係を考慮した重要度 I_c

隣接するシーンでは内容的な関連性を持つと考えたときに，重要なシーンの前後にあるシーンも重要であると推測できる．そのため，シーン c_j の重要度を計算するときに，UPRISE [2] と同様に前後 δ 枚のシーンの重要度を考慮に入れた I_c の算出式を以下のように求める．

$$I_c(c_j, \delta, \varepsilon_1, \varepsilon_2) = \sum_{k=\gamma-\delta}^{k=\gamma+\delta} I_t(c_j) * E(j-\gamma, \varepsilon_1, \varepsilon_2) \tag{11}$$

ただし，

$$E(x, \varepsilon_1, \varepsilon_2) = \begin{cases} e^{\varepsilon_1 x} & x < 0, \\ e^{-\varepsilon_2 x} & x \geqq 0. \end{cases} \tag{12}$$

とし，ε は前後のシーンの影響度を決めるパラメータで，ε が小さいほど受ける影響が大きい．

- 講演スライド資料の性質を考慮した重要度 I_k

講演は研究内容を紹介するものとして考えたときに，利用者が最も知りたい情報は「研究の着目点」や「提案アプローチ」などが挙げられる．そこで，このようなキーポイントを示す単語の集合 $K = \{k_1, k_2, \cdots, k_Y\}$ を用意し，スライド s_i の重要度算出式 I_k を下記のように提案する．

$$I_k(s_i, \alpha) = \begin{cases} \alpha * I_t(s_i) & \exists t \in \textit{the title of } s_i \wedge t \in K, \\ I_t(s_i) & \textit{otherwise}. \end{cases} \tag{13}$$

ただし，α はこのようなシーンをどれだけ重要視するかを調節するためのパラメータである．

3.8 ダイジェストの自動作成

以下では，与えられた時間 D に納まるように，キーワードに合致する複数の講演動画アーカイブスから，ダイジェストを自動作成する手法について述べる．

3.8.1 各講演への時間配分

本研究の目的はキーワードに関連する講演コンテンツの概況を利用者に提供することで，それを実現するには，限られた時間の中に，できるだけ多くの講演をダイジェストの中に納める必要がある．しかし，1つの講演がダイジェストに占める長さが短すぎると，それ内容は利用者に伝わらない可能性が大きい．そのためここでは，1つの講演コンテンツがダイジェストに占める長さは少なくともある λ_p 以上とする．そして，検索キーワードに対する適合度（$score$）を考慮して，時間配分を行なう．

まず，$score_i > 0$ となる講演コンテンツを適合度によってランキングを行なう．その結果の並び P_R を $P_R = [p_1, p_2, \cdots, p_E]$ とする．$score_i$ は i 番目の講演 $p_i (1 \leqq i \leqq E)$ の適合度を表す．

そのとき，k 番目の講演 $p_k (1 \leqq k \leqq E)$ がダイジェストに含む長さ d_k を以下のように求める．

（1） まず，与えられた時間の一部を使ってランキング上位の講演に対して均等に割り当て，

$$d_k = \begin{cases} min(D, T(p_k)) & n = 0, \\ \lambda_p & 1 \leqq k \leqq n, \\ 0 & k > n. \end{cases} \tag{14}$$

ただし，D は自動作成するダイジェストの長さで，$T(p_k)$ は k 番目の講演の長さを表す．λ_p は $\forall k (1 \leqq k \leqq E)$ に対し，$0 < \lambda_p < T(p_k)$ を満たす．そして，n は式（15）によって求める．

$$n = min(\lfloor \frac{D_1}{\lambda_p} \rfloor, E) \tag{15}$$

ここでは，D_1 は $\frac{D}{2} < D_1 \leqq D$ を満たす．

（2） 次に，残りの時間 D_2 がゼロより大きいとき，それを検索キーワードに対する適合度 $(score_i)$ に基づき，ランキング上位のコンテンツにより多くの時間を割り当てる．

つまり $D_2 = D - \sum_{i=1}^{E} d_i > 0$ のとき，Δd_k を式（16）で求め，$d_k = min((d_k + \Delta d_k), T(p_k))$ とする．

$$\Delta d_k = \frac{D_2 * score_k}{S} \tag{16}$$

ただし，$S = \sum_{i=1}^{n} score_i$ である．

（3） さらに上述の処理を経ても剰余時間が λ_p 以上となる場合，それを残りの講演コンテンツに均等に割り当てる．

つまり，$D_3 = D - \sum_{i=1}^{n} d_i > \lambda_p$ のとき，$m = min(\lfloor \frac{D_3}{\lambda_p} \rfloor, E-n)$ とする．$m > 0$ のとき，$n < k \leqq m+n$ となる p_k に対し，$d_k = \lfloor \frac{D_3}{m} \rfloor$ とする．

3.8.2 重要シーンの抽出

以下では，割り当てた時間の長さ d が $d \geqq \lambda_p$ となる講演コンテンツ p から，重要シーンを抽出する手法を提案する．

（1） まず，講演 p のタイトルに対応するシーンを c_0 とし，それ以外のシーンを重要度によってランキングする．その結果の並び R_C を $R_C = [c_1, c_2, \cdots, c_N]$ とする．また，1つの講演コンテンツの中には複数のシーンを含む．より多くのシーンを利用者に見せるために，定数 λ'_s を用意し，1枚のシーンがダイジェストに占める時間の上限とする．そのとき，シーン $c_k (0 \leqq k \leqq N)$ がダイジェストに占める長さを l_k で表し，l_k を下記のように定義する．

$$l_k = \begin{cases} min(5000msec, T(c_0)) & k = 0, \\ min((d - \sum_{i=1}^{k-1} l_i), T(c_k), \lambda'_s) & k > 0. \end{cases} \tag{17}$$

ただし，$T(c_k)$ はシーン c_k が講演動画における時間長である．

（2） そのとき，$d = d - \sum_{i=1}^{k} l_i > 0$ となる講演 p に含むシーンに対し，$l_k = \Delta l_k + l_k$ とし，Δl_k を以下のように求める．

$$\Delta l_k = \begin{cases} 0 & k = 0, \\ min((d - \sum_{i=1}^{k-1} \Delta l_i), (T(c_k) - l_k)) & k > 0. \end{cases} \tag{18}$$

（3） 最後に，タイトル以外のシーンを全て抽出されても講演 p に割り当てた時間を全て使い切ることができなければ，剰余の時間を全てタイトルシーンに加える．

3.8.3 ダイジェストの自動作成

今まで説明したように，講演 p_i がダイジェストに含む長さ d_i を決め，そして，$d_i \geqq \lambda_p$ となった講演 p_i にあるシーンがダイジェストに含む長さを決めてきた．これらの結果を元に，ダイジェストを自動生成する．まず $d_i \geqq \lambda_p$ となったランキング上位の講演から，シーンがダイジェストに占める時間長 l_j が $l_j > 0$ となったシーンを元の講演動画における出現順に従って並べる．さらに，もし各講演に対して，時間配分を行なう際に余りを生じた場合，適合度の高い講演から抽出するシーンの長さ l_j は $l_j < \lambda_S$ で，かつ $l_j < T(c_j)$ となるシーンに対し，$l_j = \Delta l'_j + l_j$ とする．$\Delta l'_j$ は

$$\Delta l'_j = min((T(c_j) - l_j), (\lambda_s - l_j), \Delta D) \tag{19}$$

ただし，$\Delta D = D - \sum_{k=1}^{E} d_k$ とし，λ_s は $\lambda_s \leqq \lambda'_s$ を満たす定数である．

4. ダイジェスト用シーンの抽出に関する実験

4. 節では，3. 節で提案した提案した単語重み付け手法と重要シーン算出式の有効性を確認するために，実験を行なった．

表 1 実験データ

抽出した単語の種類	3278
動画の平均時間長	801.70 秒
シーンの平均時間長	32.90 秒
講演あたりの平均スライド数	23.8
スライドにおける平均単語数	34

4.1 実験の概要

4.1.1 MPMeister

本研究では，プレゼンテーションコンテンツを自動生成するシステム MPMeister [10] を用いることで作成された講演動画コンテンツを実験データとして，使用している．MPMeister ではプレゼンテーションで使用されたスライドに含むテキスト情報とそれを提示したときの時間情報が XML 形式の MPEG-7 ファイルによって記述される．

また，テキストに対する形態素解析では，日本語形態素解析システム Sen [11] を使用している．

4.1.2 実験データ

今回では，MPMeister によって生成した３０件の講演コンテンツを実験データとして使用する．その概要を下の表 1 で示す．

4.1.3 実 験 方 法

実験の進め方として，まず，ダイジェストの時間配分に関するパラメータを $D = 3$ 分，$\lambda_p = 30$ 秒，$\lambda_s = 15$ 秒，$\lambda'_s = 40$ 秒，$D_1 = 2$ 分 に固定し，２種類の単語の重み付け手法及び４種類のシーンの重要度算出式に対し，比較を行なう．また，正解集合については，４名の被験者がキーワード検索によって得られた関連講演コンテンツから，それぞれ合計 $\frac{D}{\lambda_s} = 12$ 枚 のシーンを選出し，そのうち二人以上に選ばれたシーンの集合を用いた．

この条件で以下の実験を行なった．

- ２種類の単語重み付け手法 w_p と w_s を使い，重要度算出式 I_t を利用し，ダイジェストを自動作成し，比較を行なった．
- w_p を用いた 4 種類の重要度算出式（I_t, I_d, I_c, I_k）について，比較を行なった．

また評価においては，生成したダイジェストを正解集合と比較し，それぞれの F-尺度 ($F\text{-}measure$：式 (20)) を算出する．

$$F\text{-}measure = \frac{2 * precision * recall}{precision + recall} \tag{20}$$

ただし，$Digest$ は自動生成されたダイジェストにあるシーンの集合で，$Examinee$ は被験者によって選ばれたシーンの集合を表している．$precision$ 及び $recall$ は下記の式によって求める．

$$precision = \frac{|Digest \cap Examinee|}{|Digest|} \tag{21}$$

$$recall = \frac{|Digest \cap Examinee|}{|Examinee|} \tag{22}$$

4.2 単語重み付け手法に関する実験

4.2.1 キーワード検索の結果

キーワード「ストレージ」,「ストリーム」を用いて，検索した結果と上述のパラメータを用いたときの，ダイジェスト時間

表 2　キーワード「ストレージ」の実験結果

順位	適合度 (正規化済み)	長さ (秒)	正解シーン	時間配分 (秒)
1	1.0	1420	3	53
2	0.569	993	1	43
3	0.530	1460	2	42
4	0.506	1226	2	42
5	0.250	954	1	0
6	0.194	614	1	0
合計		6667	10	180

表 3　キーワード「ストリーム」の実験結果

順位	適合度 (正規化済み)	長さ (秒)	正解シーン	時間配分 (秒)
1	1.0	592	3	63
2	0.393	692	3	43
3	0.245	922	3	38
4	0.196	820	2	36
5	0.100	954	2	0
合計		3980	13	180

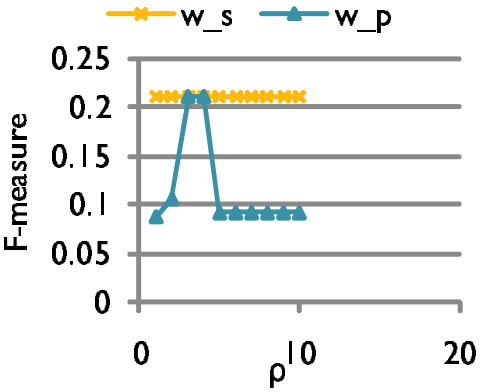

図 4　「ストレージ」を用いた時の実験結果

配分の結果を表 2，表 3 で示す．

このように，$D_1 = 2$ 分，$\lambda_p = 30$ 秒 のときでは，キーワード「ストレージ」に対する 6 件の検索結果のうち，4 件しかダイジェストに含まない．また，キーワード「ストリーム」では，5 件の検索結果のうち，4 件だけ 0 秒以上の時間が割り当てられた．

また，被験者によって選出された正解シーンの分布から，ある講演にある正解シーンの枚数の枚数は，その講演がキーワードに対する関連度に比例しないことがわかった．

4.2.2　実験結果と考察

2 つの重み付け手法（w_p と w_s）を比較するために，両方の算出式を I_t に代入したときに，生成したダイジェストの F 尺度を算出し，その結果を図 4，図 5 で示す．

w_p のパラメータ ρ を 0 から 10 まで変化させ，w_s と比較を行なった．全体からみたとき w_p が w_s を超えることができなかった．その理由として，講演で使用されるスライドの数が少ないため，isf(Inverse Slide Frequency) を単語の特定性を評価

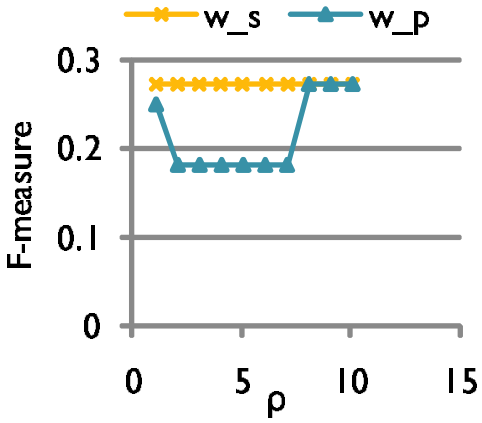

図 5　「ストリーム」を用いた時の実験結果

する指標として，十分でない可能性があることが挙げられる．

4.3　シーン重要度算出式に関する実験と考察

4.3.1　前　　提

ここでは w_s から算出した I_t を I_c，I_d，I_k に用いている．また，4 種類のシーン重要度算出式を比較するために，予備実験によってパラメータの調整を行なって，各算出式における最もよい結果をまとめ，比較を行なった．

各式におけるパラメータが $\theta = 0.5$，$\delta = 2$，$\varepsilon_1 = 5.0$，$\varepsilon_2 = 5.0$，$\alpha = 2.0$ となっている．

そして I_k で用いる特徴語集合 K は $K =$ { 背景，目的，アプローチ，提案 } である．

4.3.2　シーン重要度算出式に関する実験と考察

4 種類のシーン重要度算出式に関する実験の結果を図 6 で示す．

- 講演資料の構成パターンを考慮した算出式 I_k が最もよい結果となった．その理由として，講演のスライド資料では多くの場合，説明をわかりやすくするために「研究目的」や「アプローチ」などの単語が使用されている．

  また利用者が短時間で講演コンテンツの内容を把握するために，知りたい情報が比較的に限られているためであると考えられ，講演における特徴的題目を重視する I_k は有効である．
- I_d では，提示時間を考慮することで，結果が改善された場合があるものの，単語の重みのみ考慮した I_t よりも悪くなる場合もある．そのため，キーワードの種類を増やして再実験を行なうことが今後の課題となった．
- シーンの前後関係を考慮した I_c では，2 種類のキーワードを用いた両方の実験とも算出式 I_t と同様な結果になった．そのため，講演コンテンツの場合では，隣接するシーンの重要度を考慮した効果が極めて少ないことがわかった．

4.4　自動作成したダイジェストについて

自動作成したダイジェストでは，映像として多少不自然な部分があるが，講演資料スライドにない情報が多く含まれる．特に，要点だけ列挙するスライドに対する説明の動画の必要性を強く感じる．その反面，スライドにあるテキストをそのまま読み上げる場面もあり，このようなシーンのどう対処するかにつ

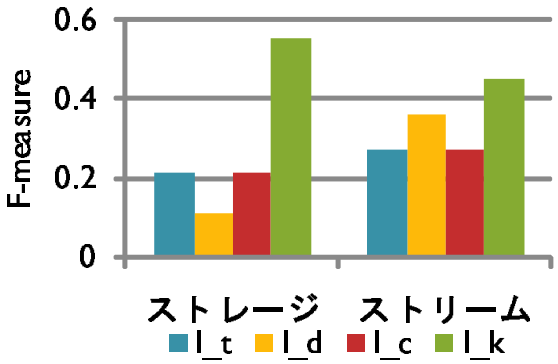

図 6　４種類のシーン重要度算出式の比較

いては，今後の課題となる．

5.　まとめと今後の課題

本稿では，講演動画アーカイブスに着目して，キーワードに合致する複数の講演コンテンツから重要と思われるシーンを抽出し，ダイジェストを自動作成する手法を提案した．提案手法ではまず講演動画アーカイブに蓄積された全講演コンテンツの中から，プレゼンテーションスライド中に含まれた検索キーワードの出現頻度に基づき，キーワードに適合する講演コンテンツをランク付けする．そして，上位のランクの講演コンテンツから重要なシーンを抽出するために，講演コンテンツにおける単語の重みについて，２種類の重み付け手法を提案し，次に重要シーンを抽出するために，４種類のシーン重要度算出式を提案した．最後に，算出したシーンの重要度及び検索キーワードに対するランキングの情報から，与えられた時間内の各講演コンテンツの時間配分を決定する手法を提案した．評価実験では，各算出式におけるパラメータの調整を行なったあと，比較を行ない，有効性を確認した．特に算出式 I_k では平均 F 尺度 0.55 で最もよい結果となった．

今後の課題として，まず今回では２種類のキーワードを用いて実験を行なったが，今後さらにキーワードの種類と講演コンテンツを増やし，実験を行なっていく必要がある．また，講演動画アーカイブを予め何らかの条件で講演を絞り込んでから，提案手法を適用するような拡張についても今後検討する必要がある．さらに，シーン重要度算出の改良が挙げられる．提案した４種類の重要度算出式ともに，スライドに含む単語の重みを利用しているが，その問題点として，スライドに出現する単語数が増えると重要度が上がることが挙げられる．その一方，文字数の少ない，図式しか含まないスライドでは重要度が低い結果になってしまう．そのため，スライドに出現する全単語の頻度や，スライドに含む図式の考慮，そして講演に対応する論文と音声情報なども考慮に入れたいと考えている．

謝　　辞

本研究の一部は，独立行政法人科学技術振興機構戦略的創造研究推進事業 CREST，および文部科学省科学研究費補助金特定領域研究 (#19024028) の助成により行なわれました．

また，本研究を進めるにあたり，筑波大学北川博之先生，森嶋厚行先生および北川研究室の皆様から貴重な実験用データをご提供いただき，この場を借りて深くお礼を申し上げます．

文　　献

[1] 日本データベース学会. DBSJ Archives. http://www.dbsj.org/Japanese/Archives/archivesIndex.html.

[2] Haruo Yokota, Takashi Kobayashi, Taichi Muraki, and Satoshi 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] 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. 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. 第 19 回電気通信学会データ工学ワープショップ (DEWS2008) 論文集, pp. E4–1, 2008.

[4] Hui Yang, Lakha Chaisorn, Yunlong Zhao, Shi-Yong Neo, and Tat-Seng Chua. VideoQA: question answering on news video. In *the 11th ACM international conference on Multimedia*, pp. 632–641. ACM, 2003.

[5] F.N. Bezerra and E.Lima. Low cost soccer video summaries based on visual rhythm. In *the 8th ACM international workshop on Multimedia information retrieval*, pp. 71–78. ACM, 2006.

[6] 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-I, No. 3, pp. 715–726, 3 2005.

[7] 藤井敦, 伊藤克亘, 秋葉友良, 石川徹也. 音声言語データの構造化に基づく講演発表の自動要約. 話し言葉の科学と工学ワークショップ講演予稿集, pp. 173–177, 2001.

[8] 堀智織, 古井貞熙. 講演音声の自動要約の試み. 話し言葉の科学と工学ワークショップ講演予稿集, pp. 165–171, 2001.

[9] Gerald Salton, editor. *Automatic text processing*. Addison-Wesley Longman Publishing Co., Inc., 1988.

[10] Ricoh Japan. MPMeister II. http://www.ricoh.co.jp/mpmeister/.

[11] 形態素解析システム Sen. http://www.mlab.im.dendai.ac.jp/ yamada/ir/MorphologicalAnalyzer/Sen.html.