講義講演ビデオの重要シーン抽出によるダイジェスト自動作成

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‡ 横田治夫††

†東京工業大学 工学部情報工学科 ††東京工業大学 学術国際情報センター
‡(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

1 はじめに

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た動画を格納し，E ーラーニング等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] のようなシステム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，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，音声情報，レーザーポインタ情報等を用いる UPRISE[2] を提案し，その効果を示してきた．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，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．

スポーツ中継やニュース映像に対して，ダイジェストを作成する研究は既に多くされているが，得点シーンやアンカマンシーン等のない講義講演コンテンツにそれらの手法をそのまま適用することはできない．一方，講義講演コンテンツでは，UPRISE のようにスライド中の単語情報等を利用することができる．そこで，本稿では，講義講演ビデオのダイジェスト自動作成を目的に，スライド中の単語出現状況等を使った重要シーンの抽出によるダイジェスト作成を考える．また，MPMeister によって作成した実際の講義コンテンツを使って，提案手法の有効性を検証する．

2 提案手法

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 1 つの講義 L_l とする．と同時に，システムは全講義で用いられた全スライド $S = \{s_1, s_2, \ldots, s_N\}$ 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 $s_i = [v_{i_1}, v_{i_2}, \ldots, v_{i_m}]$ として表現される．ここで，全スライド S 中の全単語の集合を $W = \{w_1, w_2, \ldots, w_M\}$ とすると，当然 $v_{i_\alpha} \in W$ 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 $L_l = [c_{l_1}, c_{l_2}, \ldots, c_{l_n}]$ 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 $c_{l_k} = [v_{i_1}, v_{i_2}, \ldots, v_{i_m}]$ 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，異るシーン中に複数回現れることに注意する必要がある．

2.2 シーンの重要度算出

以下，講義 L_l のダイジェストを作成するために，「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 [3] の元に，シーン中の単語の出現状況に着目して，重要と思われるシーンの自動抽出を試みる．そのために，以下のようなシーンの重要度算出式を提案する．

2.2.1 スライド構造の考慮

シーンをその中のスライド中の単語の並びとしただけでは，スライドの構造情報を利用できない．そこで，単語 v_{i_α} がスライドのどのようなレベルに出現するかを以下の関数 $p(v_{i_\alpha})$ によって表現することにする．

$$p(v_{i_\alpha}) = \begin{cases} \rho_t & v_{i_\alpha}\text{がタイトルに出現} \\ \rho_{b_l} & v_{i_\alpha}\text{が本文のインデントレベル } b_l\text{に出現} \\ 0 & v_{i_\alpha}\text{が出現しない} \end{cases}$$

これに基づき，以下のようなスライド構造を考慮した重要度算出式 I_p を提案する．

$$I_p(c_{l_k}) = \sum_{w_x \in c_{l_k}} \Big(\sum_{c_y \in L_l} \Big(\sum_{v_z \in c_y, v_z = w_x} p(v_z) \Big) \Big)$$

2.2.2 提示時間の考慮

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 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．

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

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ーである．

2.2.3 単語の出現頻度の考慮

シーン中の単語数が多ければ，I_p，I_d の値は大きくなるが，出現単語の種類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．

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

Important Scenes Extraction for Automatic Digest Generation from a Presentation Video

LE Hieu Hanh†, Thitiporn LERTRUSDACHAKUL‡and Haruo YOKOTA††
†Dept. of Computer Science, Faculty of Engineering, Tokyo Institute of Technology, 152-8552, Tokyo, Japan
‡Application Lab, Software R&D Group, Ricoh Company, Ltd., 112-0002, Tokyo, Japan
††Global Scientific Information and Computing Center, Tokyo Institute of Technology, 152-8550,Tokyo Japan
†hanhlh@de.cs.titech.ac.jp
‡thitiporn.lertrusdachakul@nts.ricoh.co.jp
††yokota@cs.titech.ac.jp

ここで、$app(c_{l_k}, w_x)$ 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.

2.2.4 単語の特定性の考慮

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.

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

なお, 上の算出式での I_{pl} は以下のように定義される.

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

2.2.5 シーンの順序の考慮

ある単語が前後のシーンにも出現すれば, 重要度が高まると考えられることから, UPRISE と同様に講義の文脈情報を考慮することも有効であると期待できる. そのようなシーンの順序を考慮した算出式 I_{dc} を提案する.

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

δ は前後何枚まで影響を与えるかを決めるパラメーターで, $E(j-\gamma, \varepsilon_1, \varepsilon_2)$ 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.

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

2.2.6 特定性とシーンの順序の考慮

最後に, 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.

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

2.3 ダイジェスト自動作成

算出式を利用してダイジェストを作成する流れを以下に示す.

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 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 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.

以上の前処理を行った後, 各算出式を適用し, 平均値以上のシーンをダイジェストのシーン候補とする. 次に, 各シーン候補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. さらに, 一般的には各シーンの最初の部分の方が重要度が高いことが多いため, 各シーン候補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. 最後に, 各シーン候補の開始時間, 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.

3 評価実験

実際に行われた 2007 年度のデータベースに関する講義録画を用いて, 被験者による評価を行った. 予め 6 名の各被験者にダイジェスト向きのシーンを正解セットとして選んでもらい, 次に各算出式に基づくダイジェストシーン候補とその正解セットの比較を行った. なお, 本評価では, 予備実験の結果から, $\rho_t = 5$, $\forall b_l\ \rho_{b_l} = 1$, $\theta = 1$, $\delta = 3$, $\varepsilon_1 = 5$, $\varepsilon_2 = 0.5$ とし,「練習問題」のシーンは性質が大きく異るため, 対象から外した. 以下の図に, 正解セットと各算出式によるシーン候補との食い違いを示す.

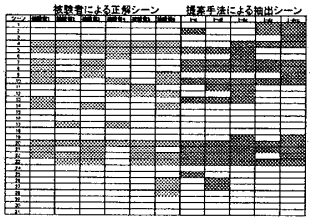

図:被験者と提案手法によるダイジェストシーン候補

図から分かるように正解セットは被験者により異るが, 各算出式によって抽出されたシーンは, 各被験者の共通部分を概ね含んでいる. また実際に作成されたダイジェストを被験者に見てもらったところ, ダイジェスト性があるという評価を得た. なお, 更に詳しい評価結果は別報 [4] を参照されたい.

4 まとめと今後の課題

講義講演ビデオのダイジェスト作成の第一歩として, 各シーンの重要度を算出する複数の式の提案と, 実際の講義を用いた評価結果を報告した. 今後は, 更に講演者の音声情報やオントロジーを利用する方法等を検討して行きたいと考えている.

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.

参考文献

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

[2] 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.

[3] H.P 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.

[4] 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, 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 回電子情報通信学会データ工学ワークショップ (DEWS2008) 論文集, 2008.