

## 科学研究費助成事業 研究成果報告書

平成 26 年 6 月 27 日現在

機関番号：53301

研究種目：基盤研究(C)

研究期間：2011～2013

課題番号：23501192

研究課題名(和文) 知識を活用した講義ビデオ検索システムの開発

研究課題名(英文) Development of the lecture video retrieval system based on knowledge

研究代表者

金寺 登 (KANEDERA, Noboru)

石川工業高等専門学校・その他部局等・教授

研究者番号：50194931

交付決定額(研究期間全体)：(直接経費) 3,900,000円、(間接経費) 1,170,000円

研究成果の概要(和文)：本研究では、知識を活用して、抽象化表現したキーワードにも対応できる講義ビデオ検索システムを開発した。クライアント側をWebアプリケーションとして構築することで、PCだけでなくモバイル端末等からもシステムが利用可能となり、授業ビデオを用いた学習効率が格段に向上した。検索方法としてキーワード補完とコンテンツ補完を導入した。コンテンツ補完において、最初のサブピックが重要であることがわかった。キーワード補完では下位概念知識の利用が効果的であった。

研究成果の概要(英文)：We have developed a search system of the class video. The system analyzes the video or analyzes the video using sound recognition using sound recognition using automatic speech recognition. This system enables us to search the scene that needed in the video. By configuring the client system as a web application, the system is now available from not only PC but also the mobile terminals. Therefore, the learning efficiency using class video increased. The method of spoken document retrieval was examined using the contents complement and keyword expansion. It was found that a beginning subtopic is useful as topic information in the contents complement. The expansion of the retrieval keyword by using the subordinate concept was effective. The subword was effective because a retrieving keyword was not contained in target in many cases. Moreover, the method using the contents complement, keyword expansion, and subword was better than the individual use.

研究分野：総合領域

科研費の分科・細目：科学教育・教育工学

キーワード：教材情報システム ビデオ教材 サブピック検索 検索キーワード補完 音声認識 講義ビデオ

1. 研究開始当初の背景

ブラウザでキーワードを指定し、インターネット上から必要な情報を得ようとしても、必要のない情報に埋もれて必要な情報を取り出すことが困難であることが多い。また、必要な情報があるページが見つかったとしても、ページ中の必要箇所を見つけるのに時間がかかる。特にビデオの場合には、目的の場面の抽出に時間を要する。膨大なビデオ情報から必要なシーンを効率よく検索することは非常に重要な問題となっている。予習や復習のために講義ビデオや各種コンテンツを効率よく活用するためには、必要な教材や教材中の場所を敏速に検索できることが望ましい。

2. 研究の目的

知識を活用して、抽象化表現したキーワードにも対応できる講義ビデオ検索システムの開発を本研究の目的とする。知識には、辞書や、Web から得られる上位・下位概念知識といった統計的な情報に加え、ルールベースのヒュリスティックモデル(人間が持つ知識をルール化したモデル)を利用する。

知識を活用することによって、講義ビデオ検索精度がどの程度、向上するかを明らかにする。予備実験を実施したところ、理想的な検索キーワードと比較して、実際のキーワードでは検索性能が悪くなる。例えば、講義内容を書き起こしたテキストに対して、3個のキーワードを用いたとき、理想的な検索キーワードでは MRR(平均逆数順位) が 0.988 であるのに対し、実際のキーワードでは MRR が 0.685 に低下する。平均逆数順位が 0.5 の場合、その逆数の 2 位に正解があることに相当する。また、音声認識を用いたときには、理想的な検索キーワードでは MRR が 0.766 であるのに対し、実際のキーワードでは MRR が 0.501 に低下する。よって、実際のキーワードに対して、知識を活用することによって、検索精度や満足度がどの程度、向上するかを明らかにする。

増大するビデオ情報をより効率良く検索することは、深刻な課題となっている。知識を活用することによって、指定されたキーワードが直接発話されていない場合にも対応できる点が本研究の特色である。特に知識中のどの概念が重要かを定量的に調査する点が独創的な点である。また、その成果をシステムとして、公開することで、社会的にも貢献できる。

3. 研究の方法

(1) キーワード補完

テキスト検索では、検索キーワードを補完する方法が利用されることがある。本研究では、音声情報検索に対して、検索キーワードの補完がどの程度有効であるかを調査する。

図1のように検索キーワードを補完する方法を情報源で分類すると、辞書を用いる方法、web 等の情報を用いる方法、併用する方法な

どが考えられる。辞書を用いる方法は一般用語を多く含むが専門用語を含んでいないことが多い。逆に、web 等の情報を用いる方法は、一般用語を含まないこともあるが、専門用語を含む可能性が高い。

概念で分類すると、上位概念、下位概念、同義語などが考えられる。これらの概念を検索することができる日本語 Wordnet を用いて調査することとした。

講義ビデオシーン(サブトピック)検索方法を図2に示す。検索対象講義ビデオをシーン分割しておき、シーン毎に TF-IDF などのインデックスを求めておく。検索時には、指定されたキーワード及びキーワードから辞書や Wordnet, Web などの知識から連想される連想キーワードから検索ベクトルを求める。この検索ベクトルとシーン毎のインデックスを比較し、類似シーンを検索する。なお、ベクトル間の類似度には Pivoted Document Length Normalization を用いた。



図1. キーワード補完

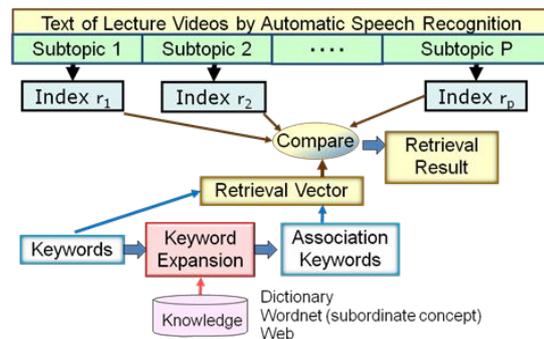


図2. シーン検索方法

(2) コンテンツ補完

図3のように、授業や講演などを検索する場合、検索したい該当シーンに話題の情報が含まれないことがある。幸い講演等では、最初に概要を述べることが多い。また講演等の最後にはまとめとして、講演全体の要約が含まれることがある。よって講演等の最初のシーン、最後のシーン、前後のシーンの情報は、話題情報として有用と考えられる。そこで  $i$  番

目の講演中の  $j$  番目のシーンのインデックス  $V_{i,j}$  を以下のように定義する。

$$V_{i,j} = (1 - \beta_s - 2\beta_a - \beta_e)T_{i,j} + \beta_s T_{i,1} + \beta_a (T_{i,j-1} + T_{i,j+1}) + \beta_e T_{i,N_i}$$

ここで  $T_{i,j}$  は  $i$  番目の講演中の  $j$  番目のシーンの TF-IDF ベクトル,  $N_i$  は  $i$  番目の講演のシーン数,  $\beta_s, \beta_a, \beta_e$  はそれぞれ最初のシーン, 隣接シーン, 最後のシーンに対応する係数である。

このようなコンテンツ補完により, 講演内での共通認識を各シーンのインデックスに反映させることができる。

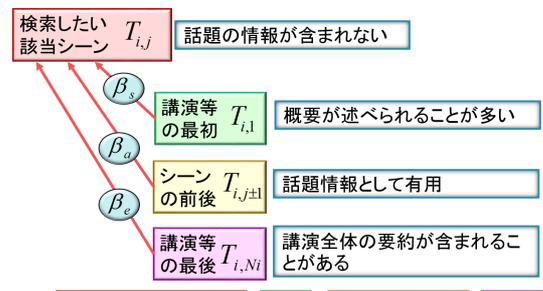


図3. コンテンツ補完

### (3) サブワード検索

正しく音声認識されるとは限らないので, 未知語対策として知られているサブワードを用いる方法についても調査した。図4のようにサブワードとして triphone を用いた。

事前準備として, 音響モデル(triphone モデル)間の confusion matrix を作成した。分布間の距離には, Bhattachayya distance を用いた。また, 検索対象の講義音声を生声認識し, シーン毎の triphone 系列を求めておく。

キーワードが未知語の場合, サブワード連続 DP によるシーン検索を行った。キーワードの triphone 系列とシーン毎の triphone 系列を confusion matrix を用いて照合した。

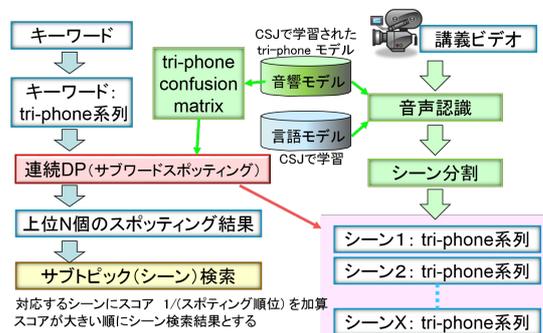


図4. サブワード検索

サブワード連続 DP の結果, スポットティング距離の小さい  $N$  個(本実験では 500 個) を抽出する。  $N$  個を距離の小さい順に順位づけし, 対応するシーンに  $1/\text{順位}$  をスコアとして加える。スコアの大きいシーンをシーン検索結果とした。

## 4. 研究成果

### (1) 評価条件

検索対象は CSJ に含まれる学会講演及び模擬講演 2702 講演とし, 30 文毎に分割したシーンの検索を行った。

各種パラメータを決定するために NTCIR-9 で提供されたドライラン 39 クエリを使用した。また, NTCIR-9 で提供された単語単位の音声認識テキストを利用した。

評価検索実験は NTCIR-10 で提供されたドライラン 120 クエリを使用した。また, NTCIR-10 で提供された単語単位の音声認識テキストを利用した。

### (2) 評価指標

評価尺度には各クエリの平均適合率をクエリ数で平均した MAP (Mean Average Precision) を用いる。クエリ  $q$  に対し, 全正解文書数を  $|R_q|$ , 検索された上位 1000 件に含まれる正解文書を  $r_1, r_2, \dots, r_M$  ( $M \leq |R_q|$ ) としたとき, 平均適合率  $AveP_q$  及び MAP は, 以下のように計算できる。

$$AveP_q = \frac{1}{|R_q|} \sum_{k=1}^M \frac{k}{rank(r_k)}$$

$$MAP = \frac{1}{|Q|} \sum_{q \in Q} AveP_q$$

ここで,  $rank(r_k)$  は正解文書  $r_k$  の順位,  $|Q|$  はクエリ総数である。

### (3) 結果

キーワード補完には, 日本語 Wordnet の下位概念と NDK 基本語辞書を用いた。クエリ中の名詞等の自立語に対応する検索ベクトル要素に一定値  $\alpha$  を設定し, 補完された単語に対応する検索ベクトル要素に  $(1 - \alpha) / (\text{補完単語数})$  を設定する。

Wordnet を用いたキーワード補完による検索結果を図5に示す。ベースラインは, 補完を用いない場合の結果である。この結果より, 連想キーワードに下位概念を用いた方が, 上位概念や下位概念と上位概念の両方を用いるよりも検索性能が高いことがわかった。上位概念や下位概念と上位概念の両方を用いた場合は, 検索に関係のない単語が連想されたためと考えられる。

コンテンツ補完とキーワード補完による検索結果を表1及び図6に示す。表中のテキストは、講義音声の書き起こしテキストを検索対象とした場合で、ASRは講義音声を音声認識したものを検索対象とした場合の結果である。コンテンツ補完のみでは、検索性能が劣化しているが、コンテンツ補完とキーワード拡張を併用することによって検索性能が向上した。

NDK 基本語辞書のみを用いたキーワード補完(ASR)における $\alpha$ の最適値は、0.45であった。日本語 Wordnetのみを用いたキーワード補完における $\alpha$ の最適値は、0.40であった。NDK 基本語辞書と日本語 Wordnetの両方を用いたキーワード補完における $\alpha$ の最適値は、0.44であった。

コンテンツ補完(ASR)における $\beta_s, \beta_a, \beta_e$ の最適値は、 $\beta_s = 0.16$ ,  $\beta_a = 0.001$ ,  $\beta_e = 0.0001$ であった。これより最初のシーンが話題情報として有用であることがわかる。コンテンツ補完とキーワード補完を併用した場合の各係数の最適値は、 $\alpha = 0.45$ ,  $\beta_s = 0.15$ ,  $\beta_a = 0.005$ ,  $\beta_e = 0.0005$ であった。

キーワード拡張、コンテンツ補完、サブワードによる検索結果を図7に示す。サブワード検索とコンテンツ補完を併用した場合にも検索性能が向上した。最も検索性能が高いのはキーワード拡張、コンテンツ補完、サブワードすべてを併用した場合であった。

表1. コンテンツ補完とキーワード補完による検索結果

補完方法	MAP	
	テキスト	ASR
ベースライン	0.1161	0.0460
コンテンツ補完	0.1183	0.0450
キーワード補完(NDK 基本語辞書)	0.1160	0.0457
キーワード補完(wordnet)	0.1159	0.0467
キーワード補完(wordnet+NDK)	0.1157	0.0458
コンテンツ補完+キーワード補完(wordnet+NDK)	0.1175	0.0468

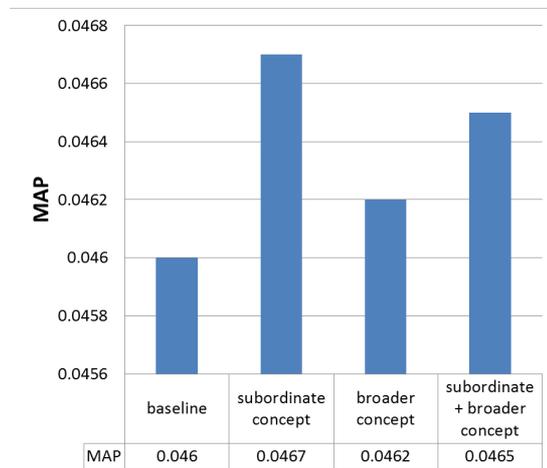


図5. キーワード補完(Wordnet)による検索結果

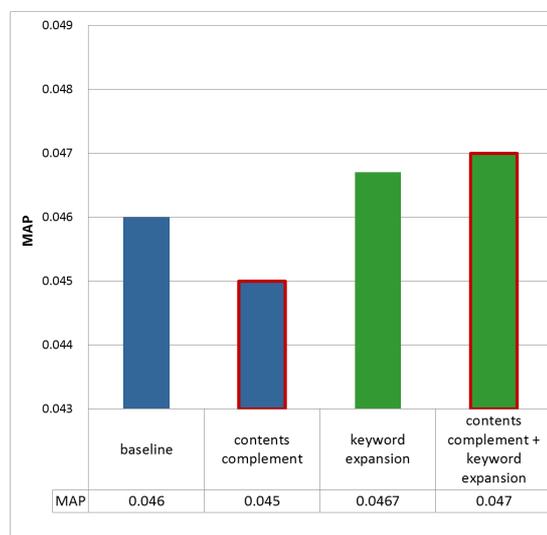


図6. キーワード拡張, コンテンツ補完による検索結果

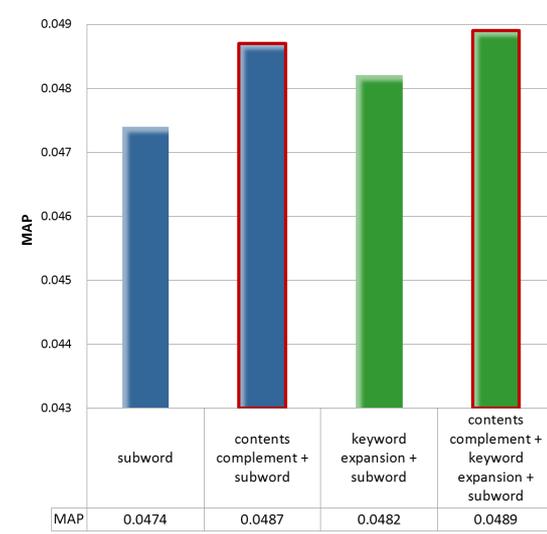


図7. キーワード拡張, コンテンツ補完, サブワードによる検索結果

(4) ビデオ検索システムの開発と公開

システムの負荷分散とシステムを利用する端末の機種依存を考慮して、図8のようにサーバ・クライアント方式を採用した。クライアントはPCやモバイル端末等から利用可能にするため、Webアプリケーションとして作成を行いブラウザから利用可能とした。システムの管理である授業の登録やビデオの登録は、全てブラウザ上で行えるようになっている。

ビデオ検索は、図9のように「年度」「授業」を指定することで検索可能であり、「日付」「検索ワード」でさらに絞り込むことができるものにした。結果として出力された検索結果に対応するビデオは図10のようにブラウザ上で再生できるものとなっている。また、ユーザアカウントを登録することで個人メモをビデオごとに保存できるようにした。

本研究の成果である授業ビデオ検索システムは、<http://sail.i.shikawa-nct.ac.jp/whale/>より、自由にダウンロードし利用できるように公開している。

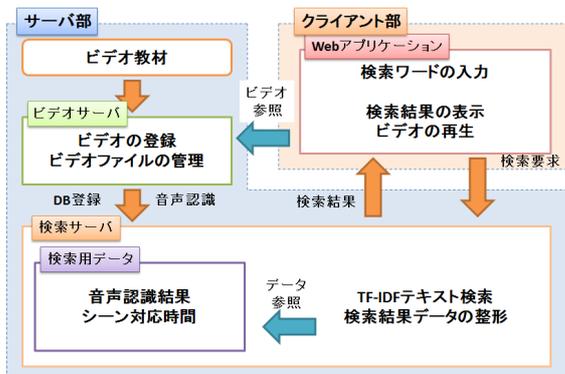


図8 . システム構成



図9 . 検索画面

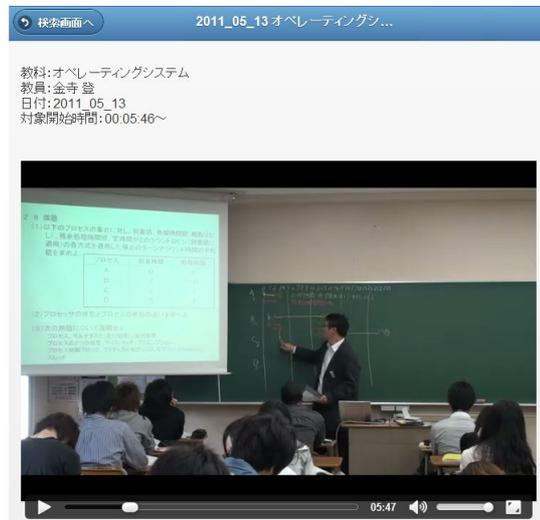


図10 . 講義ビデオ再生画面

5 . 主な発表論文等

〔雑誌論文〕(計3件)

Makoto Koshino, Yuki Kojima, Noboru Kanedera, Development and Evaluation of Educational Materials for Embedded Systems to Increase the Learning Motivation, US-China Education Review A, 査読有, Vol. 3, No. 5, 2013, 305 - 313

越田 和基, 金寺 登, 授業ビデオ検索システムの開発, 石川高専紀要, 査読有, No.45, 2013, 53 - 58

Kimhuoch Pek, Takayuki Arai, Noboru Kanedera, Voice activity detection in noise using modulation spectrum of speech: Investigation of speech frequency and modulation frequency ranges, Acoustical Science and Technology, 査読有, Vol.33, No.1, 2012, 33 - 44

〔学会発表〕(計9件)

金寺 登, 越田和基, 授業ビデオ検索システムの開発, 日本音響学会 2013 年秋季研究発表会, 71-72, 2013.9.26, 豊橋

Noboru Kanedera, Spoken Document Retrieval by contents complement and keyword expansion using subordinate concept for NTCIR SpokenDoc, Proceedings of the 10th NTCIR Conference, 634 - 637, 2013.6.20, 東京

金寺 登, キーワード補完とコンテンツ補完による音声情報検索, 日本音響学会 2013 年春季研究発表会, 105 - 106, 2013.3.15, 東京

Makoto Koshino, Yuki Kojima,  
Noboru Kanedera, Development  
and Evaluation of Educational  
Materials for Embedded Systems  
to Increase the Learning  
Motivation、International  
Symposium on Advances in  
Technology Education 査読有  
2012.9.20、熊本

越田和基、金寺 登、授業ビデオ検  
索システムの改良、平成24年度  
電気関係学会北陸支部連合大会、  
2012.9.2、富山

金寺 登、意欲喚起と自学自習支援  
による専門教育の充実、平成23年  
度国立高等専門学校教員顕彰 文部  
科学大臣賞 受賞講演、2012.8.28、  
東京

Youji Yamada, Noboru Kanedera,  
Ryotaro Komura, Syuichi Okano,  
Improvement of Learning Process  
Through Project-Based Learning  
and Repetitive Learning、  
International Symposium on  
Advances in Technology  
Education、査読有、2011.9.28、  
Singapore

越田和基、金寺 登、授業ビデオ検索  
システムの開発、平成23年度 電気  
関係学会北陸支部連合大会、  
2011.9.17、福井

Yoshikazu Kawayoke, Yasuhiro  
Inazumi, Noboru Kanedera, An  
Objective Assessment Scale of  
Visual Likeness for  
Pseudo-Representation Image  
Coding, The Fifth International  
Workshop on Image Media  
Quality and its Applications、査読  
有、2011.5.5、京都

〔その他〕

ホームページ

<http://sail.i.ishikawa-nct.ac.jp/whale/>

## 6. 研究組織

### (1) 研究代表者

金寺 登 (KANEDERA, Noboru)

石川工業高等専門学校・電子情報工学  
科・教授

研究者番号：50194931