社団法人 情報処理学会 研究報告
IPSJ SIG Technical Report

2007－NL－179 (15)
2007－SLP－ 66 (15)
2007／5／25

講義音声自動要約のための重要文手がかり表現の自動抽出

藤井 康寿 † 北岡 教英 ‡ 中川 聖一 †

† 豊橋技術科学大学情報工学系 〒441-8580 愛知県豊橋市天伯町雲雀ヶ丘1の1
‡ 名古屋大学大学院情報科学研究科メディア科学専攻 〒464-8603 名古屋市千種区不老町
E-mail : fujii@slp.ics.tut.ac.jp, kitaoka@nagoya-u.jp, nakagawa@slp.ics.tut.ac.jp

概要

我々は，大学院における講義音声を対象として，書き起こし，要約およびセグメンテーション／インデキシングといった一連の自動処理について研究している．本稿では，重要文抽出に基づく自動要約のための新しい素性の抽出方法を提案する．本稿で提案する方法は，重要文の手掛かり表現（Cue Phrase for important sentences: CP）を自動抽出するものである．本手法により，CP 自動抽出は CRF を用いた文へのラベリング問題として定式化される．本手法により抽出された素性単独の自動要約結果は precision において，書き起こしを使用した場合には 0.603，音声認識結果を使用した場合には 0.556 であった．本手法により抽出された素性と，我々が従来から用いている素性（表層的言語情報として頻出単語，スライド中の頻出単語，Term Frequency (TF)，韻律情報としてパワー，発話時間長を使用）を組み合わせた結果，音声認識結果を用いた場合において κ 値で 0.380，F 値で 0.539，*Rouge*-4 では 0.709 という比較的良い要約結果を得た．

キーワード 講義音声，音声自動要約，重要文抽出，重要文手がかり表現

Automatic Extraction of Cue Phrases for Important Sentences for Automatic Lecture Speech Summarization

† Yasuhisa FUJII ‡ Norihide KITAOKA † Seiichi NAKAGAWA
† Department of Information and Computer Sciences, Toyohashi University of Technology
1-1, Hibarigaoka, Tempaku-cho, Toyohashi 441-8580, Japan
‡ Department of Media Science Graduate School of Information Science, Nagoya University
E-mail : fujii@slp.ics.tut.ac.jp, kitaoka@sp.m.is.nagoya-u.ac.jp,nakagawa@slp.ics.tut.ac.jp

Abstract

We investigates automac summarization of spoken class-room lectures. This paper presents a novel method for sentence extraction-based automatic speech summarization. We propose a technique that extracts "cue phrases for important sentences (CPs)" that often appear in important sentences. We formulate CP extraction as a labeling problem to word sequences and use Conditional Random Fields (CRF) for labeling. Automatic summarization using CP extraction results as features yields precisions of 0.603 and 0.556 when using manual transcriptions and Automatic Speech Recognition (ASR) results, respectively. Combining the features derived from the CPs and traditional features proposed by us (including repeated words, words repeated in a slide text, and Term Frequency (TF), which are surface linguistic information, and speech power and duration, which are prosodic features), we obtained better summarization performance with a κ-value of 0.380, a F-measure of 0.539, and a *Rouge*-4 of 0.709 for speech recognition transcriptions.

key words Classroom lecture speech, Automatic speech summarization, Sentence extraction, Cue phrase for important sentences

1 はじめに

ネットワークの容量増加に伴い，講義や講演をビデオ録画し，自宅からでも容易に学習や復習が可能になってきている．もし，インデキシングや要約された音声データを使用することが可能になれば，音声データの利便性はずっと高まると予想できる．そのため近年，音声要約や自動インデキシング，セグメンテーションの研究が注目を集めている [1, 2].

我々も講義音声の有効利用に関する研究として，講義音声の認識・要約・インデックス化について研究を行っている [2]．本稿では，その中でも特に要約について報告する．

我々はこれまで，韻律情報と表層的言語情報に基づく重要文抽出による要約について研究を行ってきた [3, 4, 5]．この時，文とはポーズ長を基に音声を自動的

に区切ったものを人手もしくは音声認識器によって書き起こしたもので，抽出した文に対応する音声波形を結合することで，容易に要約音声を作成可能である．重要文抽出に使用する韻律情報としては，話速，パワーが有効であり，表層的言語情報としては，頻出単語，TF，スライド中の頻出単語が有効であった．これらの素性を組み合わせることで，κ 値や F 値においてそれぞれを単独で組み合わせるよりも良い結果を得ることができたが，その結果は人手による要約からは若干見劣りするものであった．また，自動要約は「要点の保存」について 10 人に聞いた聴取実験においても，人手による要約に若干劣るものであった [6]．

本稿では，重要文抽出の高精度化を目指し，まず，自動要約に有効な素性として，重要文の手掛かり表現 (CP) を自動抽出する手法を提案する．そして，CP 抽出結果より得られる素性と，従来から使用している素性を組み合わせた場合の要約結果を示す．CP を，重要文中に良く出現し，非重要文中には殆んど出現しない表現と定義すると，CP は重要文の手掛かり表現として有用であり，CP の情報を利用することで，重要文抽出に基づく自動要約を改善することが期待できる．本稿において，CP 抽出は文へのラベリング問題として定式化され，ラベリングには Conditional Random Fields (CRF) [7] を使用する．

2 重要文抽出法

本節では，重要文抽出法について説明する．重要文抽出は 図 1 に示される手順で行われ，文集合から重要度の高い文集合が抽出される．まず，文より文の重要性を決定付ける素性が抽出され，その後，その素性に基づいてその文が重要文かそうでないかが判定される．

素性抽出と，抽出した素性の組み合わせによる重要文分類法を 2.1 節と 2.2 節で紹介する．

2.1 素性抽出

文からは表層的言語情報と韻律情報の 2 種類の素性を抽出する．以降それぞれについて説明する．

2.1.1 表層的言語情報

表層的言語情報とは，手掛かり語（句），頻出する単語，文の位置などの言語の表層的な情報で，先行研究 [5] において以下に示すような素性が比較的重要文抽出には有用であった．

- **頻出単語** ($WORD_{rpt}$)
  文中に頻出する単語．形態素解析ソフト茶筅 [8] を用いて文を形態素解析し，講義中に出現する名詞の出現回数を数えることで，頻出単語を検出する．頻出単語を 2 語以上含む文は重要であるとして抽出される．
- **スライド中の頻出単語** ($WORD_{srpt}$)
  近年，大学における講義はパワーポイントのスライドを利用して行われることが多いため，多くの場合スライド中に出現する単語の情報は利用可能である．今回は，スライド中に頻出する単語を 2 語以上含む文は重要であるとした．
- **Term Frequency** (TF)
  不要語やフィラーを除去し，カウントする品詞を名詞のみとして，TF ベースの簡易要約器 Posum [9] を利用して要約を作成する．この TF による要約をベースラインとする．

2.1.2 韻律情報

韻律情報の使用により，要約精度を高めることが期待できる．本稿では，先行研究 [3] において比較的有効であった $F0$，パワー，発話時間長を韻律情報として使用した．

- **パワー** ($Power$)
  パワーの平均が高いものから順に重要であるとする．
- **発話時間長** ($Duration$)
  発話時間長が長いものから順に重要であるとする．
- **話速** ($Speed$)
  話速の速いものから順に重要であるとする [6]．

2.2 重要文分類法

文 i のスコアは以下のように計算される：

$$Score(S_i) = \boldsymbol{w}\boldsymbol{x} + b\ , \qquad (1)$$

ここで，$\boldsymbol{w}$ は各素性の重みを並べたベクトル，$\boldsymbol{x}$ は各素性の値を並べたベクトルであり，b はバイアスを表す．本稿において，$\boldsymbol{w}$ は SVM [10] により，$\boldsymbol{x}$ をマージン最大化に基づいて分離するように推定される[*1]．$\boldsymbol{x}$ の各要素の値には，対応する素性が頻出単語 ($WORD_{rpt}$) などのように単語を含む／含まないといった 2 値的な素性の場合には 2 値を使用し，TF やパワーのように連続値を持つ場合にはそれを平均 0 分散 1 に正規化し

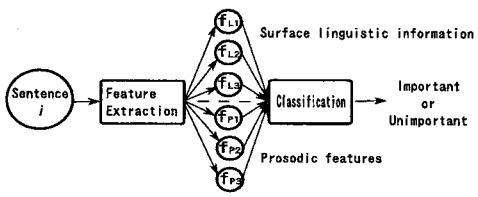

図 1 重要文抽出法の概要

[*1] 本稿においては，カーネルには線形カーネルを使用する．理論上はさらに複雑なカーネルを使用することができるが，学習データの不足により，過学習を起こす．

て連続値を使用する．

3 重要文手掛かり表現の自動抽出

もし，重要文中に良く出現するが，非重要文中には殆んど出現しないような表現を知ることが出きれば，それは重要文抽出の良い手がかりになるはずである．そのような表現を重要文手掛かり表現（Cue Phrases for important sentences: CP）と呼ぶ．CP を上記のように定義した場合，CP を含む文は重要文である確率が高いといえる．しかし，CP は講義や話者によって異なっていることが予想され，単語（形態素）の系列として一般的に CP を定義することは難しい．つまり，ある講義における CP がわかったとしても，それを他の講義に適用するのは難しいと推察される．そこで本稿では，CP を直接定義するのではなく，CP を形成するルール（パターン）は話者や講義によって限定されないという仮説のもとに，このルールを推定することによって，CP を抽出するアプローチを提案する．

CP を形成するルールは CRF で学習する．CRF によって学習するものは，学習データに対して付与する CP ラベル列の規則である．CRF を用いることで，CP 自動抽出を文に対するラベリング問題と考えることができる．

3.1 Conditional Random Fields: CRF

Conditional Random Fields（CRF）[7] は，セグメンテーションやラベリングを行うための確率モデルを構築するためのフレームワークであり，入力系列 $\boldsymbol{x}$ に対する出力系列 $\boldsymbol{y}$ の確率 $P(\boldsymbol{y}|\boldsymbol{x})$ を直接最大化するように学習を行う識別モデルである．CRF において，確率 $P(\boldsymbol{y}|\boldsymbol{x})$ は以下のように記述される．

$$P(\boldsymbol{y}|\boldsymbol{x}) = \frac{\exp\langle \boldsymbol{\Theta}, \boldsymbol{\Phi}(\boldsymbol{x}, \boldsymbol{y})\rangle}{\sum_{y \in Y} \exp\langle \boldsymbol{\Theta}, \boldsymbol{\Phi}(\boldsymbol{x}, \boldsymbol{y})\rangle} \qquad (2)$$

ここで，$\boldsymbol{\Theta}$ は各素性の重要度を表すベクトルであり，$\boldsymbol{\Phi}(\boldsymbol{x}, \boldsymbol{y})$ は各素性が成立する個数をベクトルとして並べたものである．

3.2 CP 抽出方法

CP は，図 2 に示される手順で抽出する．まず，学習データから学習用の CP が抽出し，それに基づき学習データに対してラベル付けを行う．次に，CRF によってラベル付けのルールを学習する．学習したルールによってテストデータにラベル付けを行うことで CP を抽出する．以降それぞれの手順について説明する．

3.2.1 学習データにおける CP 抽出とラベル付け

まず，学習データに対してラベル付けを行うために，学習データにおける CP を抽出する．CP を抽出するために，学習データから CP の候補となる表現のリストを作成する．CP の候補となる表現とは，学習データ

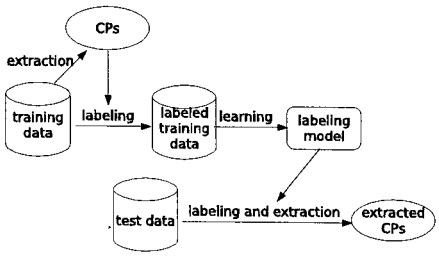

図 2 重要文手がかり表現抽出手順

表 1 ラベル付けの規則

ラベル	意味
0	非重要語
1	CP の先頭語
2	CP の中間語
3	CP の終端語
-1	CP 中の非重要語（スキップ語）

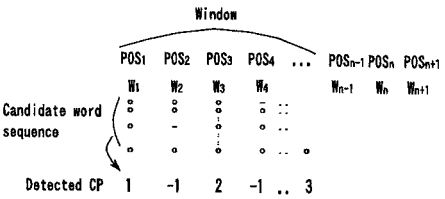

図 3 学習データにおけるラベル付け

（文毎に重要／非重要情報を持つ）の各文を形態素解析した結果中の，連続する 8 つの形態素内（単語窓内）の 3 形態素以上 8 形態素以下の全ての組合わせ（連続していなくても良い）である．この時，形態素どうしが隣接していない場合にはその間を正規表現の '.*' で表す．形態素解析には『茶筅』を用いた．これらの CP 候補に対して，学習データにおいて以下の CP の条件に合致するものを CP とみなして抽出する．ある表現 e が条件をみたすとは，① 学習データの重要文中に出現する回数 $C_I(e)$ が Th_N 回以上，② その表現を含む文が重要文である条件付確率 $P_I(e) = C_I(e)/(C_I(e) + C_N(e))$ が Th_R 以上，となる場合である．ここで，$C_I(e)$ および $C_N(e)$ はそれぞれ，表現 e を含む重要文数および非重要文数である．

次に，学習データ中の CP に対してラベル付けを行う．ラベル付けは，表 1 に基づいて行う．表における CP 中の非重要語とは，表現中の '.*'（0 個以上の形態素）にあたる．図 3 に学習データにおけるラベル付けを図示する．

3.2.2 CRF の学習

ラベル付けされた学習データについて CRF の学習を行う．図 4 は，CRF の素性関数を示す．遷移素性として隣接する状態，観測素性として自身と前後の形態素を用いる．

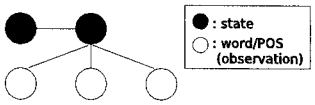

図 4 CRF の素性関数

3.2.3 テストデータにおける CP 抽出と CP 抽出による素性

テストデータに対して学習した CRF を適用することで，文の CP 箇所にラベル付けすることができる．各文に対し，CP がラベル付けされたか否かを CP 抽出による素性（CP）とする．

3.2.4 CRF への入力列

CRF への入力列として，形態素列と品詞列が利用可能である．重要文の手がかり表現について考えたとき，3.2.1 節の方法で CP を抽出，ラベル付けすると，形態素列を用いた場合には話者やドメインが強力に限定されてしまう可能性がある．また，品詞だけを用いた場合には，一般化されすぎて手がかり表現を表現できるか疑問である．そこで，ドメインを限定せず，また，表現能力を失わないように形態素と品詞を混合することを考える．形態素を用いた場合に，ドメインに依存するのは名詞であると推測できるので，形態素と品詞の混合列は，名詞については品詞を使用し，その他については形態素を使用するものとする．本稿では，比較のために形態素列，品詞列，そして形態素列と品詞列の混合の 3 種類について CP 抽出実験を行う．比較結果は 4.4 節に示す．

4 実験

4.1 実験条件

今回の実験には，4 名によって行われた合計 8 個の講義音声データを使用した．これらの講義は，本学大学院で開講されている音声言語処理，マルチモーダルインタフェース，パターン認識，および自然言語処理に関連するものである．表 2 (a) は講義音声試料の特性を示す．各講義はおよそ 70 分で，文に換算すると約 1000 文である．ここで文とは，200ms 以上の無音区間で自動的に区切られた各区間を指す．要約率は文数ベースで 25% とした．

全ての音声データは，人手および音声認識器によって書き起こされる．音声認識器には SPOJUS [11] を用いた．音響モデルには音節単位の HMM，言語モデルにはトライグラムを使用して 2 パスで認識する．音響モデルおよび言語モデルは CSJ コーパス [12] から学習した．本稿で使用する講義音声の認識性能は，単語認識精度において 31.0% から 57.5% であった．詳細を表 2 (b) に示す．

CP 抽出および SVM の学習の際には，4-fold のクロスバリデーションを行う．ある話者をテストデータとするとき，残り 3 人の学習データを用いて学習を行う．

3.2.1 節における Th_N と Th_R は経験的にそれぞれ 10 と 0.75 とした．

4.2 要約の正解

要約の正解を作成するために，人手による要約を作成する．講義は大学院修士向けのもので専門性が高いため，対象講義の分野に造詣が深い 6 人に要約を依頼した．各被験者には，文単位で設定要約率（今回は 25%）と等しくなるように各文に重要かそうでないかのラベル付けを行ってもらった．

要約の正解には，各個人による要約のばらつきを吸収するために **man3/6** を使用する．man3/6 とは，6 人中 3 人以上が重要と判断した文は重要であるとして 6 人の要約から作成する重要文集合である．man3/6 は，各被験者間の差異を軽減することがわかっている [4]．

表 2 に示す要約の目標値は，ある被験者と その被験者を除いた **man3/5**（その被験者を除いた 3 人以上が重要と判断）との一致度の平均である．

4.3 評価尺度

評価尺度には，κ 値 [13]，Precision，F 値，そして $Rouge$-N [14] を用いる．それぞれ次のように定義される．

- **Precision, Recall**

$$Precision = \frac{|M \cap H|}{|M|}, \quad Recall = \frac{|M \cap H|}{|H|}.$$

ここで，H と M はそれぞれ人手による抽出文と自動要約による抽出文集合である．本稿では，素性の評価に Precision を用いる．

- **F 値**

F 値は Precision と Recall の調和平均として定義される：

表 2 音声試料諸元

(a) 講義音声の特性と人間による要約結果

Lecture	Duration	No. of Sent.	Target value of summarry		
			κ	F	$Rouge$-4
SN-1	67'56"	742	0.462	0.595	0.632
SN-2	54'59"	719	0.491	0.613	0.633
NK-1	65'49"	680	0.474	0.599	0.547
NK-2	71'14"	1099	0.450	0.579	0.638
TN-1	69'28"	582	0.493	0.617	0.618
TN-2	78'30"	648	0.320	0.447	0.527
TA-1	70'02"	1749	0.454	0.586	0.615
TA-2	65'23"	1571	0.477	0.605	0.592
average	67'55"	974	0.453	0.580	0.600

(b) 音声認識精度

Lecture	Accuracy [%]	Correct [%]
SN-1	47.4	55.6
SN-2	31.0	37.0
NK-1	54.9	60.8
NK-2	50.7	58.9
TN-1	48.8	54.8
TN-2	45.0	55.2
TA-1	57.1	61.4
TA-2	57.5	62.5

$$F\text{-}measure = \frac{2 \times Precision \times Recall}{Precision + Recall}$$

- **κ 値** [13]
  κ 値とは，2 者の判定の一致度を，偶然の一致を考慮して調整した指標であり，以下のように定義される：

$$\kappa = \frac{P(A) - P(E)}{1 - P(E)}$$

$$P(A) = \frac{\text{重要文の一致数} + \text{非重要文の一致数}}{\text{文の総数}}$$

$$P(E) = \text{重要文の偶然の一致率} + \text{非重要文の偶然の一致率}$$

$$\text{重要文の偶然の一致率} = \frac{A\text{が重要と判定した文}}{\text{文の総数}} \times \frac{B\text{が重要と判定した文}}{\text{文の総数}}$$

$$\text{非重要文の偶然の一致率} = \frac{A\text{が重要でないと判定した文}}{\text{文の総数}} \times \frac{B\text{が重要でないと判定した文}}{\text{文の総数}}$$

- ***Rouge*-N** [14]
  Rouge-N は，N-gram の再現率を表す．*Rouge*-N は各被験者の要約結果をそのまま使用することができるので，*Rouge*-N のみ **man3/6** ではなく，各被験者の要約を正解として用いる．また，本稿では N=4 とし，*Rouge*-4 を使用する．

$$ROUGE - N = \frac{\sum_{S \in \{Ref-Summaries\}} \sum_{gram_n \in S} Count_{match}(gram_n)}{\sum_{S \in \{Ref-Summaries\}} \sum_{gram_n \in S} Count(gram_n)}$$

4.4 CRF への入力列の比較

CRF への入力列として，形態素，品詞，形態素と品詞の混合列をそれぞれ使用した場合の CP 抽出実験を行い，それぞれについてラベルが付与された文を重要文とみなして文抽出を行った結果を表 3 に示す．値は音声認識結果を使用した時のもので，各講義の結果の平均値である．表より，文抽出の精度は形態素を使用した時が一番高いが，形態素を使用した場合には極端に抽出文数が少ないため，素性としてはあまり有効ではないといえる．品詞と，品詞と形態素の混合を比べた場合，品詞と形態素の混合の方が Precision が高く，抽出文数に大差ないため，CRF への入力列としては形態素と品詞の混合列を使用するのが良いといえる．

表 3 各入力列を使用した場合の文抽出結果

Input Seq.	Word	POS	Word & POS
# of Ext. Sent.	7.4	96.8	51.6
Precision	0.609	0.529	0.556

4.5 抽出された CP 例

図 5 に，本手法により抽出された表現例を示す．音声認識結果を使用する場合の方が，一般的に表現中に名詞が多く含まれることが多い．これは，音声認識誤りにより助詞をうまく扱えないことによると推測できる．

人手による書き起こし使用
- 例えば * と * は
- *noun* という * *noun*
- こういう * *noun* は * の * だ

音声認識結果使用
- は * *noun* * だ
- *noun* * *noun noun* は
- *noun* * の *noun* * は *noun*

図 5 抽出された CP 例（* は任意の文字列）

表 4 CP 抽出結果に基づく重要文抽出結果（Precision）

(a) Using transcriptions by human

Spoken Lecture	Extracted	Important	Prec.
SN-1	112	58	0.518
SN-2	85	48	0.565
NK-1	46	34	0.739
NK-2	24	17	0.708
TN-1	81	40	0.494
TN-2	95	44	0.463
TA-1	57	33	0.579
TA-2	37	28	0.757
average	67.1	37.8	0.603

(a) Using ASR transcriptions

Spoken Lecture	Extracted	Important	Prec.
SN-1	82	40	0.488
SN-2	59	32	0.542
NK-1	43	26	0.605
NK-2	29	20	0.690
TN-1	47	23	0.489
TN-2	85	37	0.435
TA-1	33	18	0.545
TA-2	35	23	0.657
average	51.6	27.4	0.556

4.6 CP 抽出結果に基づいた重要文抽出結果

CP 抽出結果に基づいて文抽出を行った結果を表 4 に示す．学習・テストデータに人手による書き起こしを使用した場合の Precision は 0.603 で，音声認識結果を使用した場合の Precision は 0.556 であった．

図 6 は，2.1 節で説明した各素性単独の要約結果と CP 抽出結果を利用した要約結果を，音声認識結果を使用した場合について比較している．各素性の値は，各講義における値を平均したものである．図より，CP 抽出結果は他のどの素性よりも精度が高く，有効な素性であることがわかる．書き起こしを使用した場合の結果についても同様の結果となった．

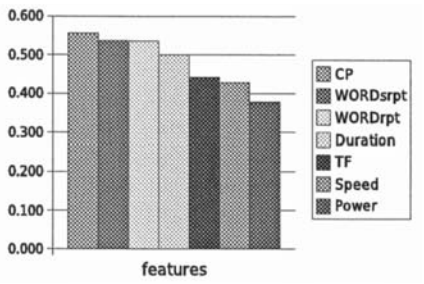

図 6 各素性単独の重要文抽出結果
（Precision，音声認識結果使用，要約率 25%）

表 5 従来素性の組み合わせによる要約結果と従来素性に CP を組み合わせた場合の要約結果（ASR）

Lecture	κ-value		Rouge-4	
	prev.	with CP	prev.	with CP
SN-1	0.294	0.294	0.733	0.733
SN-2	0.343	**0.350**	0.720	**0.729**
NK-1	0.439	**0.447**	0.735	0.735
NK-2	0.472	**0.476**	0.795	**0.797**
TN-1	0.315	**0.324**	0.650	**0.660**
TN-2	0.365	0.365	**0.677**	0.675
TA-1	0.343	**0.349**	0.620	**0.621**
TA-2	0.428	**0.431**	**0.730**	0.725
average	0.375	**0.380**	0.708	**0.709**

表 6 要約結果の比較

Trans.	Features	κ	F	*Rouge-4*
Manual	TF (baseline)	0.200	0.408	0.443
	previous features [5]	0.374	0.534	0.716
	+ new feature	0.375	0.535	0.718
ASR	TF (baseline)	0.230	0.427	0.466
	previous features [5]	0.375	0.535	0.708
	+ new feature	0.380	0.539	0.709
human		0.453	0.580	0.600

4.7 重要文抽出結果

音声認識結果を利用した場合について，2.1 節に示した従来から用いている素性のみを利用して要約した結果（prev.）と，それらに CP 抽出により得られた素性を加えた場合の要約結果（with CP）を合わせて表 5 に示す（F 値の値は κ 値と類似しているため，紙面の都合上省略）．表 5 に示すとおり，従来の素性に CP を加えることによって多くの講義において効果がみられ，各評価尺度での要約結果の平均はそれぞれ，$\kappa = 0.380$, *Rouge*-4 $= 0.709$ であった．

表 6 は，人手の書き起こしを使用した場合および音声認識結果を使用した場合におけるベースライン（TF），従来の素性のみを使用した場合，従来の素性に CP を加えた場合，そして，人手による要約の結果を示す．表より，素性を組み合わせることで，単独素性による結果であるベースラインよりも良い結果を得ることができており，また，CP を加えた要約は，従来の素性を組み合わせた要約結果を上回っていることがわかる．しかし，CP を加えた要約でも，κ 値および F 値においては人間による要約に依然およばないといえる．*Rouge*-4 において自動要約が人手による要約を上回っているが，これは自動要約が長い文を選択しやすいことによるものであると考えられる．

5 おわりに

本稿では，CRF を用いた重要文手掛かり表現（CP）の自動抽出方法を紹介し，CP 抽出による素性が自動要約に与える影響について示した．CP 抽出結果単独による要約結果は，Precision で，人手による書き起こしを使用した場合には 0.603，音声認識結果を利用した場合には 0.556 であった．また，CP 抽出により得られる素性を従来の素性と組み合わせた結果，従来の素性のみを使用した場合の結果を上回ることがわかり，κ 値では 0.380，F 値では 0.539，*Rouge*-4 では 0.709 の比較的良い要約結果を得た．

今後は，現在は文自身に含まれる手掛かり表現抽出しかできていないが，提案手法を文間の表現を抽出できるように拡張していく．また，講義音声だけでなく，講演音声についても本手法の効果を確認する．

参考文献

[1] 金寺, 隅田, 池端, 船田. ビデオ教材作成支援を目的とした講義音声によるシーン分割. 電子情報通信学会論文誌, Vol. DI, No. 5, pp. 977–984, 2005. 5.

[2] 富樫, 山口, 北岡, 中川. 講義音声の認識・要約・インデックス化の検討. 情報処理学会研究報告, SLP-62-11, 2006. 7.

[3] S. Kobayashi, N. Yoshikawa, and N. Nakagawa. Extracting summarization of lectures based on linguistic surface and prosodic information. *SSPR*, pp. 211–214, 2003. 4.

[4] 小林, 山口, 中川. 表層的言語情報と韻律情報を用いた講演音声の重要文抽出. 自然言語処理, Vol. 12, No. 6, pp. 3–24, 2005. 11.

[5] S. Togashi, M. Yamaguchi, and S. Nakagawa. Summarization of spoken lectures based on linguistic surface and prosodic information. *IEEE/ACL Workshop on Spoken Language Technology*, pp. 34–37, 2006. 12.

[6] 藤井, 山口, 北岡, 中川. 韻律・表層的言語情報に基づく重要文抽出による講義音声要約の評価. 日本音響学会講論集, 2-P-28, 2006. 9.

[7] F. Pereira J. Lafferty, A. McCallum. Conditional random fields: Probabilistic models for segmenting and labeling sequence data. In *Proceedings of the 18th International Conference on Machine Learning*, 2001.

[8] 松本裕治, 北内啓, 山下達雄, 平野善隆, 浅原 正幸松田寛. 日本語形態素解析システム『茶筌』 version 2.2.1 使用説明書, 2000.

[9] H. Mochizuki. *Automatically Text Summarizer Posum, version 1.50.2.* Japan Advanced Institute of Science and Technology http://www.tufs.ac.jp/ts/personal/motizuki/software/posumcl/, 2002.

[10] M. O. Stitson, J.A.E. Weston, A Gammerman, V.Vork, and V. Vapnik. Theory of support vector machines. Technical Report CSD-TR-96-17, Department of Computer Science, Royal Holloway College, University of London, 1996. 12.

[11] 北岡, 高橋, 中川. N-best 線形辞書検索と 1-best 近似木構造辞書探索の併用による大語彙連続音声認識. 電子情報通信学会論文誌, Vol. 87-DII, No. 3, 2004. 3.

[12] S. Furui, K. Maekawa, and H Isahara. A japanese national project on spontaneous speech corpus and processing technology. *Proc. ASR2000*, pp. 244–248, 2000.

[13] J. L. Fleiss. Measuring nominal scale agreement among many rater. *Psychological Bulletin*, Vol. 76, pp. 378–382, 1971.

[14] C. Lyn and E. Hovy. Automatic evaluation of summaries using n-gram co-occurence statistics. *the Human Language Technology Conference*, pp. 71–78, 2003.