情報処理学会論文誌 Vol. 51 No. 3 1094–1106 (Mar. 2010)

重要文抽出に基づく講義音声の自動要約

藤井康寿[†1] 山本一公[†1]
北岡教英[†2] 中川聖一[†1]

本論文では，大学院における講義音声を対象とした，重要文抽出に基づく自動要約手法を述べる．本論文ではまず，音声要約においてよく使われている Maximal Marginal Relevance（MMR）と識別器に Support Vector Machine（SVM）を用いた feature-based を比較し，feature-based の方が優れた結果を与えることを示す．次に，feature-based の改善手法に関して述べる．Feature-based の改善のために，3 つのアプローチを試みた．1 つ目は，重要文中によく出現するような重要文の手がかり表現（Cue Phrase for important sentences; CP）を自動抽出し，自動要約の素性とする手法である．CP の抽出は Conditional Random Fields（CRF）を用いてラベリング問題として定式化される．2 つ目は，人間による要約は重要文が連続しやすいという観測に基づき，重要文の連続性を考慮した要約を行う方法である．連続性を考慮するために，連続性をとらえる新たな素性を使用する．3 つ目は，冗長性を排除する枠組みの導入である．これらの 3 手法を用いることで，feature-based による要約を改善できた．

Class Lecture Summarization Based on Important Sentence Extraction

Yasuhisa Fujii,[†1] Kazumasa Yamamoto,[†1]
Norihide Kitaoka[†2] and Seiichi Nakagawa[†1]

This paper describes summarization methods based on important sentence extraction for the summarization of class-room lecture. First, we compare two summarization techniques; a Maximal Marginal Relevance and a feature-based method which uses a Support Vector Machine (SVM) as a classifier. We show that the latter is superior to the former. Second, we improve the feature-based summarizer by three different types of approaches. In the first approach, we propose a technique that extracts “cue phrases for important sentences (CPs)” that often appear in important sentences and thus can be used as a feature to the summarizer. We formulate CP extraction as a labeling problem of word sequences and use Conditional Random Fields (CRF) for labeling. The second approach presents a novel sentence extraction framework that takes into account the consecutiveness of important sentences based on the observation that important sentences tend to be extracted consecutively by human. We deal with this consecutiveness by applying this new features to a feature-based summarizer. The third approach provides a way to reduce redundancy in the summary. Experimental result shows that our method outperforms traditional sentence extraction methods using these aproaches.

1. はじめに

近年，オンラインでアクセスして使用できる講義コンテンツの量が飛躍的に増加しており，これらのコンテンツに対する音声認識の技術が研究されている[1),2)]．もし，これらのコンテンツに対して索引を付与できたり，重要な箇所のみを提示するような要約を作成することができたりすれば，これらのコンテンツの利便性はずっと高まり，より扱いやすくなる．そのため，インデキシングや音声自動要約の研究はこれまで以上に注目を集めている[3)–7)]．

要約は，抜粋型とアブストラクト型に分けることができる．抜粋型とは，要約対象全体であるドキュメントを単語や文などの単位に分割し，なんらかの基準に基づいてそれらの単語や文を抽出する方法であり，アブストラクト型は，原文の言い換えを許した自由作文型の要約である．音声の要約においては，最終的な媒体が音声である場合，高精度な音声認識の困難さや滑らかな音声生成の必要性のために，一部の語句の削除や言い換えによる要約は難しいため，抜粋型で要約することが一般的である[8)]．また，最終的な音声の自然さを考慮すると，抽出に用いる単位はある程度の長さを持っている必要があるため，抽出の単位としては文が使用されることが多い．このように，文を単位とし，重要な文を抽出することで要約を作成する方法を重要文抽出法と呼ぶ．重要文抽出による音声要約手法は，近年でも世界中で広く研究されている[4)–6),9),10)]．日本においても，「話し言葉工学」プロジェクト（CSJ）コーパスの開発によって，講演音声に対する重要文抽出型要約の研究が広く行われた[8),11)–13)]．

本論文では，日本語講義音声コンテンツコーパス CJLC[14)] に含まれる講義音声を使用し，講義音声に対する重要文抽出法を検討する．本論文で扱うような長時間の講義音声コンテン

†1 豊橋技術科学大学情報工学系
Department of Information and Computer Sciences, Toyohashi University of Technology
†2 名古屋大学大学院情報科学研究科メディア科学専攻
Department of Media Science, Graduate School of Information Science, Nagoya University

ツに対する要約の研究はこれまでなされていない．そのため，講義音声のような長時間の音声コンテンツに対して，どのような要約手法が適しているかは検討されていない．

CSJ などの学会講演発表音声（以下，単に講演音声と呼ぶ）では，あらかじめ定義した手がかり語（例：「結果をまとめると」，「という結果になりました」）が，手がかり語を含むような重要文や，その手がかり語の周辺に存在する重要文の手がかりとして使用できたが[8]，講義音声は，話者による表現も異なり，講演音声に比べて発話スタイルがよりくだけ，内容が整っていないため，あらかじめ手がかり語を定義することが困難であり，講演音声で有効であった手がかり語をそのまま講義音声の要約に適用しても効果が少ない．また，講演音声では，その中に比較的明確に現れる構造の情報（序論，本論，結論）が有効に用いられたが[12]，講義音声は講演音声に比べて 1 つのコンテンツが長く，話題が広く分布しているため，構造情報は必ずしも有効でない．さらに，限られた時間で研究成果を発表するために冗長性が少ない講演音声に対して，講義音声では発話スタイルがより自由であるため，理解を促すために同じ内容を複数回繰り返し発話することなどから生じる冗長性の排除に関して，特に考慮する必要がある．

講義音声に対して高精度な重要文抽出を実現するために，まず，音声要約で広く用いられている重要文抽出法の比較を行う．音声要約の研究においては，MMR が広く有効な手法として用いられているが[6),9),10),15)]，我々は従来から feature-based による手法を用いてきている[8),16)]．MMR と feature-based による手法を比較することで，本論文で提案している SVM を使用した feature-based 手法が，MMR よりも講義音声を要約するために適した手法であることを示す．

さらに，講義音声の feature-based による要約をより高精度化するために，3 種類の新しい素性と，それらを用いた要約手法を提案する．

1 つ目は，重要文中によく出現し，非重要文中にはほとんど出現しないような重要文の手がかり表現（Cue Phrases for important sentences; CP）のパターンを機械学習によって自動的に学習し，CP 抽出結果を新たな素性として使用する手法である．機械学習を用いて表現のパターンを自動的に学習することで，あらかじめ手がかり語を定義しておくことが難しい講義音声において手がかりとなる表現を使用することが可能となる．

2 つ目は，人間の要約において観測される重要文の連続性に着目し，これを素性として反映させることで，精度の改善を試みる手法である．重要文の連続性を考慮することで，全体の構造情報によって重要文の出現傾向を考慮することが難しい講義音声において，局所的な重要文の出現パターンを考慮することが可能となる．

3 つ目は，MMR 法に基づいた内容の重複度合いを示す素性を新たに導入することで，冗長性の排除を試みる手法である．冗長性を考慮した素性を使用することで，講義音声において対処が必須である冗長性の排除を可能にする．

本論文は以下のように構成される．2 章において，本論文で使用する音声試料と要約の正解，評価尺度について説明する．3 章では，重要文抽出法による要約における代表的な手法である MMR と feature-based について説明し，両者による要約結果を比較する．Feature-based による要約を改善するための 3 手法の提案と，それぞれの手法の評価は 4 章で行う．最後に，5 節において本論文をまとめる．

2. コーパスと評価尺度

2.1 音 声 試 料

本論文では，日本語講義音声コンテンツコーパス CJLC[14)] に含まれる 4 人の話者による 8 講義分を対象に要約実験を行う．各講義は，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，自然言語処理に関係する内容で，本学大学院において実際に実施されている講義である．**表 1** に音声試料の諸元を示す．各講義は平均 70 分の長さで，約 1,000 文からなる．ここで文とは，200 ms 以上の無音区間で自動的に区切られた各区間を指す．実験で使用する書き起こしは，人手による書き起こしについては CJLC 付属のものを使用し，音声認識結果については SPOJUS[17)] による文単位の認識結果を用いた．音響モデルにはコンテキスト独立の音節単位の HMM，言語モデルには語彙 2 万語のトライグラムを使用して 2 パスで認識した．音響モデルおよび言語モデルは CSJ コーパス[18)] から学習した．本論文で使用する講義音声の単語認識性能は，Accuracy で平均 49.1%，Correct で平均 55.8%であった[19)]．

2.2 要約の正解と目標値

CJLC には，各講義に対して 6 人の被験者が重要文抽出による要約を行ったデータが含まれている．CJLC に含まれる講義は大学院における講義であり専門性が高いため，要約を行った被験者は講義の内容を十分に理解することができる音声言語処理関係の専門家である．各被験者は，全体の 25%の文を抽出するように指示されて要約を行った．人間の要約はばらつきが多いため，個々の要約を直接正解とはせずに，3 人以上の被験者が正解であると判断した文の集合を正解文集合とした．本論文ではこれを **man3/6** と呼ぶ．man3/6 により被験者間の一致をとることで，被験者間のばらつきを吸収することが可能である[16)]．man3/6 の要約率の平均値は表 1 に示すように 26.7%であり，設定要約率である 25%と近い値になった．

自動要約の目標は，人間に近い要約を機械によって自動的に作成することであるので，自

表 1　音声試料の諸元
Table 1　Details of speech materials.

講義 ID	時間長	文数	Acc. [%]	Corr. [%]	要約率		要約の目標値（対 man3/5）		
					man3/5	man3/6	κ	F	ROUGE-4
L11M0011	67'56"	742	47.4	55.6	0.232	0.279	0.462	0.595	0.686
L11M0012	54'59"	719	31.0	37.0	0.229	0.267	0.489	0.612	0.700
L11M0031	65'49"	680	54.9	60.8	0.235	0.276	0.484	0.609	0.674
L11M0032	71'14"	1099	50.7	58.9	0.219	0.267	0.450	0.579	0.719
L11M0041	69'28"	582	48.8	54.8	0.234	0.278	0.493	0.618	0.686
L11M0042	78'30"	648	45.0	55.2	0.218	0.210	0.444	0.574	0.666
L11M0051	70'02"	1749	57.1	61.4	0.227	0.277	0.454	0.586	0.703
L11M0052	65'23"	1571	57.5	62.5	0.233	0.281	0.477	0.605	0.726
Average	67'55"	973.8	49.1	55.8	0.228	0.267	0.469	0.597	0.695

動要約の目標値は，各被験者による要約と要約の正解との一致度とすることが望ましい．しかし，man3/6 の投票には，一致度を計算したい被験者の要約が使用されているため，各被験者による要約と man3/6 との一致度の平均は正確な目標値とはならない．そこで，6 人中のある被験者による重要文集合と，その被験者を除いた 5 人の被験者のうち 3 人以上が重要と判定した文集合（**man3/5**）の間の一致度の平均値を使用する．この方法で計算した各講義における目標値を表 1 に示す．目標値は，本論文で使用する評価尺度（一致度）である κ 値，F 値，*Rouge*-4 のそれぞれについて求めた．評価尺度については次節で説明する．任意の被験者間の一致度の平均値は $\kappa = 0.387$ であり，対 man3/5 の一致度の平均値である $\kappa = 0.469$ の方が高い値を示すので，被験者間の一致度を目標値として使用するよりは，man3/5 との一致度を目標値に使用する方が厳しい目標値となる．man3/5 の要約率は，表 1 に示すように平均 22.8%となり，当然 man3/6 よりも若干要約率が高い．

2.3 評価尺度

要約の客観尺度による標準的評価手法は確立されていないが，その中でも Rouge[20] が比較的よく用いられている[21]．本論文では，我々が従来から採用している κ 統計量（κ 値）[22] および F 値に加えて，*Rouge*-N[20] を評価尺度として用いる．

- **κ 値**

  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：

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

$P(A) = A$ と B の一致率 (2)

$P(E) = A$ と B の偶然の一致率 (3)

- ***F* 値**

  F 値は Precision（適合率）と Recall（再現率）の調和平均として定義され，抽出された文集合どうしの適合率と再現率を両方考慮した一致率を示す尺度である．

$$F\text{-}measure = \frac{2 \times Precision \times Recall}{Precision + Recall}, \tag{4}$$

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

  ここで，H と M はそれぞれ人手による抽出文集合と自動要約による抽出文集合である．

- ***Rouge*-N**

  Rouge-N は，評価対象の要約の正解要約に対する N-gram の再現率を表しており，内容の保存に関して評価することが可能である．つまり，（重複して）重要な内容が複数箇所にある場合でも，どれか 1 カ所が抽出されれば高い評価値となるという利点がある．*Rouge*-N は以下のように計算される．

  Rouge-N

$$= \frac{\sum_{S \in \{Ref\text{-}Summaries\}} \sum_{gram_N \in S} Count_{match}(gram_N)}{\sum_{S \in \{Ref\text{-}Summaries\}} \sum_{gram_N \in S} Count(gram_N)}$$

  ここで，*Ref-Summaries* は正解文集合（man3/6），$gram_N \in S$ は文 S に含まれる

N グラム，$Count_{match}(gram_N)$ は $gram_N$ が評価対象の要約文集合に含まれる個数，$Count(gram_N)$ は $gram_N$ が正解文集合に含まれる個数である．本論文では N = 4 とし，*Rouge*-4 を使用する．

3. 重要文抽出法の比較

音声要約の研究においては，MMR が広く有効な手法として用いられているが[6),9),10),15)]，我々は従来から feature-based による手法を用いて音声要約を行ってきている[8),16)]．本章では，本論文で使用する MMR および feature-based の定義を示し，両手法を比較する．

3.1 Maximal Marginal Relevance

Carbonell らによって提案された Maximal Marginal Relevance（MMR）[23)] は，テキスト検索のベクトル空間モデルに基づいた要約手法であり，元々はテキスト要約を対象とした手法であるが，文献 15) や 6) に示されるように，音声要約においても有効な手法である．

MMR は，ドキュメント（本論文においては講義音声全体）との関連度と情報の新規性に基づいて抽出する文を順に決定していくことで，全体としてドキュメントとの関連が高くかつ冗長性の低い文集合を抽出することを目指す．細かな違いにより MMR にはいくつかのバリエーションが存在するが，本論文では文献 15) で定義されるものを使用する．MMR の文抽出アルゴリズムを**図 1** に示す．ステップ (1) では，使用する各変数の初期化を行う．$\boldsymbol{S}_{rk}$ は各時点までに抽出された文集合の平均ベクトルであり，零ベクトルで初期化する．$\mathbb{S}_{rk}$ および $\mathbb{S}_{nrk}$ は各時点までに抽出された文の集合および，抽出されていない文の集合であり，それぞれ空集合およびドキュメントに含まれるすべての文の集合として初期化される．$\boldsymbol{tf}_i$ は文 i に対する単語の出現頻度に基づくベクトル表現であり，本論文においては以下のように定義する．

$$\boldsymbol{tf}_i = (tf_{i,1}, tf_{i,2}, \ldots, tf_{i,w}), \tag{5}$$

$$tf_{i,w} = f_w \cdot \log\left(\frac{f_{\hat{w}}}{f_w}\right), \tag{6}$$

ここで，f_w はドキュメント中の単語 w の頻度であり，$\hat{w}$ はドキュメント中に最も出現する単語である．ドキュメント全体に分布する単語よりも，特定の箇所に集中して出現する単語の方が重要度が高いと考えられるので，$tf_{i,w}$ は Term Frequency（TF）の値をドキュメント中の最大単語頻度に基づいて修正している．$\boldsymbol{D}$ は，ドキュメントに含まれるすべての文の集合の平均ベクトルである．$tf_{i,w}$ は，単語 w の該当ドキュメント内での出現回数 f_w をもとに定義されるため，ドキュメント中の出現位置（i）にかかわらず一定の値をとる．ただし，文 i に出現しなかった単語については $tf_{i,w} = 0$ とする．ステップ (2) では，$\mathbb{S}_{nrk}$ に含まれる文集合から式 (7) を最大化する文 $\boldsymbol{S}_{max}$ を抽出する．式 (7) における Sim は 2 つのベクトル間の類似度を表し，本論文ではコサイン距離を用いる．式の第 1 項は文とドキュメントの関連度を表し，第 2 項は情報の新規性を表す．λ は，ドキュメントとの関連度と冗長性の間のトレードオフである．ステップ (3) では，抽出した $\boldsymbol{S}_{max}$ に基づいて各変数の更新を行う．ステップ (4) では，終了条件に関する判定を行い，所望の要約率に達していれば終了，そうでなければステップ (2) に戻る．

MMR は，文を TF のベクトルで表現することによって類似度計算を行うことが前提のアルゴリズムであるために，文にそれ以外の情報を持たせることが難しいという問題点がある．

実験では，図 1 のステップ (2) における λ について，0.0 ∼ 1.0 の範囲を 0.1 刻みで変化させ，講義データ全体の κ 値を最大化する値を使用した．本実験データにおいては，$\lambda = 0.6$ で最大の κ 値を示した．

ステップ (1)　初期化：

$$R = \text{抽出文数}$$
$$N = \text{ドキュメント中の文の総数}$$
$$\boldsymbol{S}_{rk} = \boldsymbol{0},$$
$$\mathbb{S}_{rk} = \phi,$$
$$\mathbb{S}_{nrk} = \{\boldsymbol{tf}_1, \boldsymbol{tf}_2, \ldots, \boldsymbol{tf}_N\},$$
$$\boldsymbol{D} = \frac{\sum_{\boldsymbol{S} \in \mathbb{S}_{nrk}} \boldsymbol{S}}{N}$$

ステップ (2)　抽出：

$$\boldsymbol{S}_{max} = \underset{\boldsymbol{S} \in \mathbb{S}_{nrk}}{\operatorname{argmax}} \{\lambda(Sim(\boldsymbol{S}, \boldsymbol{D})) - (1-\lambda)(Sim(\boldsymbol{S}, \boldsymbol{S_{rk}}))\} \tag{7}$$

ステップ (3)　更新：

$$\boldsymbol{S}_{rk} = \frac{\boldsymbol{S}_{rk} \cdot |\mathbb{S}_{rk}| + \boldsymbol{S}_{max}}{|\mathbb{S}_{rk}| + 1},$$
$$\mathbb{S}_{rk} = \mathbb{S}_{rk} \cup \{\boldsymbol{S}_{max}\},$$
$$\mathbb{S}_{nrk} = \mathbb{S}_{nrk} - \{\boldsymbol{S}_{max}\}$$

ステップ (4)　終了判定：

$|\boldsymbol{S}_{rk}| \geq R$ なら，$\mathbb{S}_{rk}$ を重要文集合として終了．そうでないならステップ (2) へ．

図 1　MMR のアルゴリズム
Fig. 1　Algorithm of MMR.

3.2 Feature-based

Feature-based による要約は，文の重要性を表す素性の抽出と抽出した素性に基づく分類

からなる．3.2.1 項において我々が従来から使用している素性[16] および素性の処理に関して説明し，3.2.3 項において使用する識別器について説明する．

3.2.1 素　　性

ニュース放送や講演音声など構造がはっきりとしている場合には，我々は従来より，言語情報と韻律情報に基づいた要約手法を提案してきた[8),16)]．

音声認識誤りが 50%程度であっても，言語情報は非常に有効な素性である[15]．本論文で使用する言語素性を以下に示す[⋆1]．

Repeated words[8),16)]：各文中に含まれる頻出単語の異なり数を素性とする．ここで，頻出単語は，フィラーや不要語を除いた品詞が名詞である単語のうち，ある一定の閾値以上の出現頻度を持つ単語と定義する．

Words in slide texts：講義はスライドに基づいて行われることが多いため，スライド中に含まれる単語は良い手がかりとなる．したがって，各文について，その文が発話されたときに使用されていたスライド中に含まれる不要語を除いた品詞が名詞である単語を含む総数を素性とする．

Term Frequency (TF)：フィラーや不要語を除いて，式 (6) の値を TF として使用する．

音声要約においては，言語情報に加えて韻律情報も使用可能であり，韻律情報を用いることで要約精度を向上させることが可能である[16]．本論文で使用する韻律情報は以下のとおりである．

Duration：各文の発話時間長を素性とする．

Power and F0：各文のパワーと F0 の平均値を素性として使用する．本論文では，パワーおよび F0 の値は ESPS[25] を用いて抽出した．

Rate of Speech：各文の話速を素性として使用する．本論文における話速とは，単位時間あたりのモーラ数である．

Pause：現在の文と直前の文との間のポーズ長および現在の文と直後の文の間のポーズ長を素性として使用する．

なお，CSJ の講演音声の要約で有効であった文の位置情報（講演の最初と最後の 10 文）は，講義音声では有効でないので素性として用いていない[16]．

3.2.2 素性の前処理

本論文における feature-based の方法は，3.2.1 項で説明した各素性の値をそのまま使用するのではなく，前処理を行う．

3.2.2.1 素性の正規化

素性の値は，講義や話者によってばらつきがあるため，正規化を行う必要がある．本手法では，講義ごとに各素性について平均 0，分散 1 となるように正規化を行う．

3.2.2.2 素性の拡張

他の素性と組み合わせることで効果を発揮する場合や，その絶対値（2 乗）に意味がある場合を考慮するために素性の拡張を行う．本手法では，素性そのままの値に加えて，2 乗した値および任意の素性間の積を素性として用いる[⋆2]．

3.2.2.3 素性の 2 値パターン化

文献 26) と同様に，本論文では素性の値を div 個の値で 2 値パターン化し，すべての素性を div 個の 2 値変数で表現する．素性の 2 値パターン化を行うことで，素性の値の大小に対して重要度が対応しないような，文のスコアに複雑に寄与する素性を扱うことができるようになる．たとえば，$div = 5$ の場合，素性の値は “00001”，“00010”，“00100”，“01000”，“10000” の 5 パターンのいずれかに分類される．実験では，$div = 5$ を用いた．

3.2.3 識　別　器

文 i のスコアは以下の式で計算する．

$$Score(S_i) = \boldsymbol{w}\boldsymbol{x} + b, \tag{8}$$

ここで，$\boldsymbol{x}$ は 3.2.1 項に示す素性の値のベクトルであり，$\boldsymbol{w}$ および b は各素性の重みおよびバイアスを表す．$\boldsymbol{w}$ および b は SVM[27] の線形カーネルによって学習する．SVM の学習には，svm^{perf} [28] を用いた．

実験では，式 (8) の $\boldsymbol{w}$，b の学習は，話者に対してオープンとなるように行った．すなわち，ある話者に対するモデルの学習には他の 3 話者のデータを使用した．評価結果は 4-fold の交差検定となる．

3.3 評 価 実 験

表 2 に MMR による要約結果と feature-based による要約結果を示す．表の抽出文数は，各講義において抽出される文数であり，総文数の 25%となる．表より，LM11M0012 および

⋆1 本論文において，単語とは形態素解析器 *ChaSen*[24] によって分割される単位を指す．

⋆2 これは，SVM の 2 次の多項式カーネルを模している．本来，3.2.3 項において 2 次の多項式カーネルを使用することで素性の拡張は代用できるが，本手法においては，3.2.2.3 に示す素性の 2 値パターン化を行っているため，2 次の多項式カーネルの使用と本手法による素性の拡張は等価にならず，2 次の多項式カーネルを使用する方がより複雑なモデルとなる．予備実験より，表 1 に示すとおり本論文で使用するデータ量が少ないため，2 次の多項式カーネルを使用すると過学習を引き起こし結果が悪くなることが分かっている．

表 2　MMR と feature-based による要約結果
Table 2　Summarization result of MMR and feature-based summarizer.

書き起こし	講義名	抽出文数	MMR			Feature-based		
			κ	F	Rouge-4	κ	F	Rouge-4
Manual	L11M0011	186	0.305	0.489	0.646	0.363	0.531	0.720
	L11M0012	180	0.369	0.532	0.614	0.336	0.507	0.664
	L11M0031	170	0.394	0.553	0.625	0.424	0.575	0.706
	L11M0032	275	0.388	0.546	0.703	0.499	0.628	0.805
	L11M0041	146	0.321	0.500	0.567	0.288	0.476	0.618
	L11M0042	162	0.261	0.430	0.605	0.376	0.537	0.682
	L11M0051	438	0.314	0.495	0.594	0.398	0.556	0.676
	L11M0052	393	0.383	0.546	0.644	0.420	0.573	0.694
	Average	243.8	0.342	0.511	0.625	0.388	0.548	0.696
ASR	L11M0011	186	0.298	0.483	0.649	0.328	0.505	0.649
	L11M0012	180	0.355	0.522	0.660	0.321	0.496	0.648
	L11M0031	170	0.333	0.508	0.618	0.409	0.564	0.685
	L11M0032	275	0.372	0.532	0.689	0.451	0.593	0.755
	L11M0041	146	0.321	0.500	0.568	0.297	0.482	0.636
	L11M0042	162	0.339	0.490	0.638	0.376	0.537	0.682
	L11M0051	438	0.322	0.499	0.563	0.377	0.541	0.668
	L11M0052	393	0.365	0.532	0.629	0.439	0.588	0.716
	Average	243.8	0.338	0.508	0.627	0.375	0.538	0.680

LM11M0041 では MMR の方が良い結果を示しているが，それ以外の講義では feature-based が MMR よりも良い性能を示しており，平均でみると明らかに feature-based が良い結果を示していることが分かる．この結果は，使用した素性はほぼ同等であるが，識別器に GMM を採用した文献 15) や言語情報のみの素性で，識別器に単純なスコアの線形和を用いた文献 6) において MMR の方が feature-based よりも良い値を示していることに反しているが，素性および識別器の構成がほぼ同じである文献 29) において feature-based の方が MMR よりも良い値を示しているのと同じ傾向である．先行研究の結果もふまえると，素性や識別器を吟味することによって，feature-based が MMR を上回ることが可能であるといえる[⋆1]．次章以降は，feature-based による要約をさらに改善する方法について検討する．

⋆1 韻律情報を使用せず，言語情報のみを用いた場合でも，feature-based による要約の性能はほとんど低下しなかった．この結果より，feature-based が MMR を上回ったのは，単に韻律情報を使用できるためだけではなく，様々な素性を効果的に組み合わせることができるためであると考えられる．

4. Feature-based の改善

本章では，3 章において講義音声の要約に対して有効性の示された feature-based による要約をより高精度化するために，3 種類の新しい素性と，それらを用いた要約手法を提案する．

4.1 重要文手がかり表現の自動抽出と feature-based による要約への適用

我々はこれまで，講演音声に関して人手で設定した重要文に現れやすい「提案いたしました」，「結果を発表します」などの手がかり語に基づく要約手法を用いてきたが，講義音声ではこのような手がかり語は使用できない．そこで，本論文ではこのような表現を自動抽出する方法を提案する．

先にも述べたように，重要文中によく出現するが，非重要文中にはほと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がかり表現（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 が分かったとしても，それを他の講義に適用するのは難しいと推察される．そこで本論文では，CP を単語列として直接定義するのではなく，話者や講義によらないより一般的なルール（パターン）が存在する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．

話者に依存しない CP を抽出するための一般的なルールを推定するために，本論文では，CP 抽出の問題を単語列に対するラベリング問題として定式化し，CP のラベリング規則（CP 抽出のルール）を CRF で学習する．CP のラベリング規則を CRF でモデル化することによって，CP を構成する規則が獲得され，CP を単語列として直接定義するよりも，話者や講義によらないより一般的な CP を抽出可能であると考えられる．

4.1.1 Conditional Random Fields: CRF

Conditional Random Fields（CRF）[30] 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．

$$P(\boldsymbol{y}|\boldsymbol{x}) = \frac{\exp\langle \boldsymbol{\Theta}, \boldsymbol{\Phi}(\boldsymbol{x}, \boldsymbol{y})\rangle}{\sum_{y \in Y} \exp\langle \boldsymbol{\Theta}, \boldsymbol{\Phi}(\boldsymbol{x}, \boldsymbol{y})\rangle} \tag{9}$$

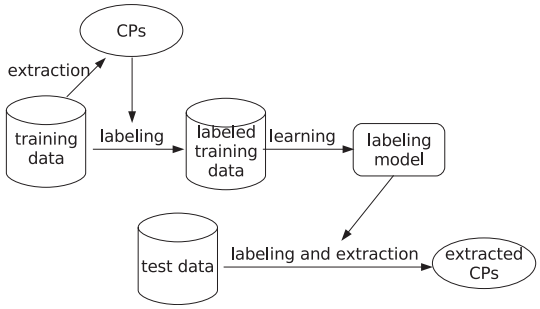


図 2　重要文手がかり表現抽出手順
Fig. 2　Procedure of cue phrases for important sentences extraction.

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 $\boldsymbol{x}$ と $\boldsymbol{y}$ の間で成立する回数をベクトルとして並べたものである．また，$\langle A, B \rangle$ は A と B の内積を示す．

4.1.2　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 を抽出し，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．

4.1.2.1　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（文ごとに重要/非重要情報を持つ）の各文中の連続する 8 つの単語窓内の 3 単語以上 8 単語以下のすべての組合せ（連続していなくてもよい）である．このとき，単語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．正規表現の '.*' に含まれる単語数は元の表現の単語数と同じである必要はない．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 と見なして抽出する．ある表現 e が条件を満たすとは，① その表現を含む重要文数 $C_I(e)$ が Th_N 以上，② その表現を含む文が重要文である条件付き確率 $P_I(e) = C_I(e)/(C_I(e) + C_N(e))$ が Th_R 以上，となる場合である．ここで，$C_N(e)$ は表現 e を含む非重要文数である．本論文では，経験的に $Th_N = 10$，$Th_R = 0.75$ とした．使用する書き起こしの違いによる影響を調査するために，学習データに人手による書き起こしを使用する場合と音声認識結果を使用する場合の両方で同じ値を用いた．

次に，こうして抽出された学習データ中の CP に対してラベル付けを行う．ラベル付けは，**表 3** 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 個以上の単語）にあたる．ラベル付けの際，1 つの単語に複数のラベルが付与される可能性がある場合には，長い CP ほど信頼度が高いとして，最長の単語数を持つ CP が付与されるようにラベリングする[*1]．**図 3** に学習データにおけるラベル付けを図示する．

表 3　ラベル付けの規則
Table 3　Labeling rule.

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）

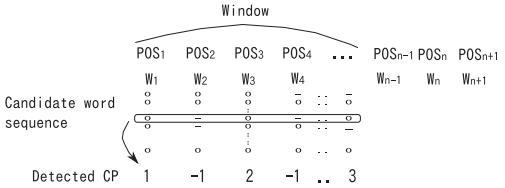


図 3　学習データにおけるラベル付け．'o' は CP 構成単語，'-' は CP 非構成単語
Fig. 3　Labeling CPs in training data. 'o' describes a word in a CP candidate, and '-' a word not in the CP candidate.

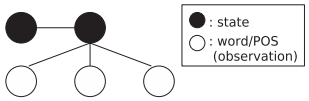


図 4　CRF のグラフィカル表現
Fig. 4　Graphical representation of CRF.

次に，こうして抽出された学習データ中の CP に対してラベル付けを行う．ラベル付けは，**表 3** 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 個以上の単語）にあたる．ラベル付けの際，1 つの単語に複数のラベルが付与される可能性がある場合には，長い CP ほど信頼度が高いとして，最長の単語数を持つ CP が付与されるようにラベリングする[*1]．**図 3** に学習データにおけるラベル付けを図示する．

4.1.2.2　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** に，CRF の素性関数を

*1 局所的に解決できない場合には，時間的に先に生じた CP（文頭に近い位置から開始した CP）を優先的にラベリングする．

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単語を用いる．CRF の学習には CRF++ [⋆1] を用いた．

CRF の学習は，3.2.3 項における SVM の学習と同様に，話者に対してオープンとなるように行った．すなわち，ある話者に対するモデルの学習には他の 3 話者のデータを使用した．評価結果は 4-fold の交差検定となる．

4.1.2.3　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個数を CP 抽出による素性（CP 素性）とする．

4.1.2.4　CRF への入力列

CRF への入力列として，単語列と品詞列が利用可能である．CRF への入力列として単語列を用いた場合には，話者や話題，ドメイン（この場合は講義内容）などが強力に限定されてしまう可能性があり，また，品詞だけを用いた場合には，一般化されすぎてしまう可能性がある．この問題のために，最もドメインに依存する品詞は「名詞」と考えられるので，品詞が「名詞」である単語については，「名詞」に置き換え，その他の単語については単語そのものを使用した系列を CRF への入力列として使用する．

4.2　Feature-based による重要文の連続性を考慮した要約

人間が行った重要文抽出結果には，重要文が連続して出現することが多い．本節では，feature-based による要約において重要文の連続性を扱う方法を提案する．

4.2.1　重要文の連続性

man3/6 における重要文数の平均は 268.0 であり（全体の約 25%），そのうち単独で出現した重要文数は 80.6 であった．また，約 70%の重要文は連続して出現していることが分かった[⋆2]．平均の重要文の連続長は 1.83 であり（各文を無作為に 1/4 抽出すると連続長の平均は 1.33），重要文の次の文が再び重要文である確率は非常に高いことが分かる．もし，この重要文の連続性をうまくとらえることができれば，要約の精度を向上させることができる可能性がある．

4.2.2　重要文の連続性を扱う素性

本論文では，重要文の連続性を扱うために，新たに 2 つの素性を使用する．

⋆1 http://chasen.org/∼taku/software/CRF++/

⋆2 CSJ の講演音声では，重要文の連続長は要約率 33%の場合で 2.93 であった（各文を無作為に 1/3 抽出した場合の期待値は 1.50）．

4.2.2.1　動 的 素 性

重要文に連続して抽出されやすい傾向があるならば，直前の文の抽出結果は現在の文の抽出結果の良い手がかりとなるはずである．この観点に基づいて，直前の文が重要文として抽出されたかどうかを動的素性として使用する．動的素性は 2 変数からなり，それぞれ直前の文が抽出されたかどうかおよび直前の文が抽出されなかったかを現し，2 値の値を持つ．文 i に対する動的素性は以下の式で表される．

$$dynamic(i) = \begin{cases} 10 & \text{if} \quad S_{i-1} \text{ is extracted.} \\ 01 & \text{otherwise.} \end{cases} \tag{10}$$

4.2.2.2　差 分 素 性

重要文が連続して抽出されるのは，その中で関連する話題を同程度の密度（素性の値）で話しているからであると考えられる．人間が複数の文にわたって関連する話題を話すとき，その話題を話し始めた文とその直前の文の素性の値は大きく変化し，同じ話題を話している間は，各文が持つ素性の値の変化は緩やかであると予想できる．この予想に基づいて，現在の文と直前の文の素性の値の差分を差分素性として用い，重要文の連続性の尺度とする．文 i からの素性 j に対する差分素性 $diff_{i,j}$ は以下のように計算される．

$$diff_{i,j} = f_j(S_i) - f_j(S_{i-1}). \tag{11}$$

差分素性で扱っているのは，重要文どうしの「内容」の類似度ではなく，素性の「値」の類似度であり，4.3.1 項に示される冗長素性が扱う類似度とはねらいが異なっている．たとえば，repeated words のような単語に基づく素性の場合においても，隣接する文どうしが同じ単語を含なくても，同数の頻出単語を含んでいれば，差分はゼロとなり，含まれる頻出単語の数に大きな違いがあれば，差分は大きくなる．

4.2.3　最適な重要文集合の同定

連続性を考慮した要約においては，素性に動的素性と差分素性を加え，式 (8) によって文の重要度を決定する．しかし，動的素性により，現在の文の重要度が直前の文の抽出/非抽出の決定に依存しているため，文ごとに独立にスコアを計算することができない．そこで，連続性を考慮した要約では，以下の式に基づき，文ごとのスコアに基づいて重要文を決定するのではなく，抽出された文集合のスコアの総和が最大となるような文集合 $\mathbb{S}_{imp}$ を抽出する．

$$\mathbb{S}_{imp} = \underset{\mathbb{S} \subset \mathbb{D}}{\operatorname{argmax}} \sum_{S \in \mathbb{S}} Score(S) \tag{12}$$

$$\text{subject to } |\mathbb{S}_{imp}| = R$$

ここで，$\mathbb{D}$ はドキュメント中の全文の集合，R は所望の要約率に基づく抽出文数，$Score(S)$ は式 (8) に基づき計算される値である．動的素性の導入によって現在の文のスコアに影響を与えるのは直前の文の抽出/非抽出の結果のみであるため，任意の要約率において式 (12) を最大化する文集合は，下記の動的計画法を解くことで一意に同定可能である．

$$g_0(i,j) = \max\begin{cases} g_0(i-1,j) \\ g_1(i-1,j) \end{cases} \tag{13}$$

$$g_1(i,j) = \max\begin{cases} g_0(i-1,j-1) + score(i|0) \\ g_1(i-1,j-1) + score(i|1), \end{cases} \tag{14}$$

ここで，i は現在の文番号で j は現在までに抽出した文数である．$g_1(i,j)$ と $g_0(i,j)$ はそれぞれ，i 番目の文を重要文として抽出した場合としない場合における，i 番目の文までに j 文を抽出した場合の最大スコアである．$score(i|1)$ と $score(i|0)$ はそれぞれ，直前の文が抽出された場合とされなかった場合における文 i を抽出するスコアであり，式 (8) によって計算される．ただし，素性には動的素性と差分素性が加えられている．全文数を I，所望の要約率を R $(= J/I)$ として，$\max(g_0(I,J), g_1(I,J))$ へのパスを求めることで，任意の要約率を満たす文集合を求めることができる．

4.3 Feature-based による冗長性を考慮した要約

講義音声では，受講者の理解を促すために同じ内容を複数回繰り返し発話することがあるなど，冗長性の排除に関して考慮することは必須である．本節では，MMR 法に基づいた素性を新たに導入することで，feature-based の要約において冗長性を排除するための手法を提案する．

4.3.1 冗長性を扱う素性

3.1 節に示す MMR 法は，文のスコアを計算する際に，その文とすでに抽出した文集合との類似度を考慮することで，冗長性を排除している．Feature-based において冗長性を考慮するために，各文と重要文集合との類似度を考える．各文が式 (5) で表現されたあるドキュメントに対して重要文集合 $\boldsymbol{imp}$ が与えられたときの，文 i に対する冗長性素性 $rdun(i)$ を下記のように計算する．

$$rdun(i) = Sim(\boldsymbol{tf}_i, Imp), \tag{15}$$

$$Imp = \begin{cases} \frac{\sum_{S \in \boldsymbol{imp} - \boldsymbol{tf}_i} S}{|\boldsymbol{imp}|} & \text{if} \quad \boldsymbol{tf}_i \in \boldsymbol{imp} \\ \frac{\sum_{S \in \boldsymbol{imp}} S}{|\boldsymbol{imp}|} & \text{otherwise,} \end{cases} \tag{16}$$

ここで Sim はコサイン類似度である．冗長性素性により，講義全体の冗長性を排除することを期待する．

4.3.2 探　　索

素性の重みを学習する時点では，式 (16) における $\boldsymbol{imp}$ は正解文集合であるが，実際に文抽出を行う際には，正解文集合は当然未知である．そのため，実際に文抽出を行う際には，$\boldsymbol{imp}$ を各時点までに抽出した文集合の平均ベクトルに置き換え，逐次更新しながら探索を行う．各時点までに抽出した文集合によって式 (15) の値は異なってくるため，式 (12) を最大化する文集合を動的計画法を用いてすべての仮説から一意に同定することはほぼ不可能である．したがって，4.2.3 項に示した連続性を考慮した要約における最適な重要文集合の同定よりも高度な探索機構が必要となる．冗長性を考慮した要約では，式 (12) を最大化する文集合の探索のために，式 (13) における $g_0(i,j)$ および $g_1(i,j)$ について，それぞれ W 個の仮説を保持することによるビームサーチを行う．実験では，ビーム幅 W を 30 とした．

4.4 評 価 実 験

4.4.1 重要文手がかり表現の自動抽出と要約への適用の評価

4.4.1.1 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．図の例からも分かるように，機械学習によって抽出された表現が，人間にとって直感的に要約に役立つと感じる例は少なかったが，実際に有効かどうかは 4.4.1.2 で検討する．

4.4.1.2 CP 抽出結果に基づく素性単独による要約結果

CP 素性単独による要約結果を**表 4** に示す．各講義において抽出された CP 数も併記してある．表 4 の実験において，重要文および非重要文中の各単語に “0” 以外のラベル（CP を構成する単語）が付与された割合を調査した結果，それぞれ，19.9%および 10.6%であった．4.1.2.3 における CP 素性は，各文に対して CP がラベル付けされた個数であるが，本項における結果は，CP が付与された文はすべて重要文であるとして文抽出を行った結果である．表 4 には κ 値も示しているが，素性は，どの程度正確に重要文を推定できるかが重要であるため，抽出結果は Precision によって評価する．人手による書き起こしを使用した場合の Precision の平均は 0.551 で，音声認識結果を使用した場合の Precision は 0.559 であった．この値は，素性単独としては非常に優れた値である[16]．たとえば，従来の素性で

人手による書き起こし使用	
・ *noun* という * *noun*	（解析ということが必要）
・ こういう * *noun* は * の * だ	（こういうよなんは人間の口だ）
・ *noun noun* の *noun* を し	（音声対話の話をし）
・ の *noun* という * は	（の感覚というのは）
音声認識結果使用	
・ *noun* * と か	（拡大とか縮小とか）
・ *noun noun* の *noun* * し て	（音声対話の利用して）
・ 次に ま *noun* * って * *noun*	（次にま特徴抽出っていうこと）
・ の * *noun* * *noun* * *noun*	（の定義先ほどの三つの活性）

図 5　抽出された CP 例（* は任意の文字列）．括弧内は CP として抽出された文字列の例
Fig. 5　Examples of extracted CPs ('*' means arbitrary words) which are extracted from the words in parenthesis.

表 4　CP 抽出結果に基づく重要文抽出結果（Precision，κ）
Table 4　Important sentence extraction results based on CP extraction (Precision, κ).

Trn.	講義名	抽出文数	Precision	κ	抽出 CP 数
Manual	L11M0011	240	0.479	0.307	402
	L11M0012	203	0.517	0.354	303
	L11M0031	94	0.628	0.287	129
	L11M0032	77	0.714	0.210	113
	L11M0041	176	0.460	0.267	286
	L11M0042	215	0.470	0.319	361
	L11M0051	160	0.525	0.143	221
	L11M0052	128	0.617	0.173	186
	Average	161.6	0.551	0.258	250.1
ASR	L11M0011	230	0.435	0.232	364
	L11M0012	165	0.497	0.282	248
	L11M0031	81	0.531	0.184	99
	L11M0032	52	0.731	0.152	59
	L11M0041	124	0.516	0.272	184
	L11M0042	173	0.509	0.330	256
	L11M0051	115	0.574	0.128	149
	L11M0052	96	0.677	0.157	131
	Average	129.5	0.559	0.217	186.3

Precision の良かったものは TF や Repeated words であるが，それらの値は，人手による書き起こしを使用した場合で 0.566，0.532，音声認識結果を使用した場合で 0.556，0.548 である．この結果より，CP 素性を新たな素性として加えることで，要約の精度を向上させることができる可能性がある．CP 素性をその他の素性と組み合わせた場合の効果については 4.4.1.3 で検証する．

表 5　CP 素性，連続性を考慮した素性，および冗長性を考慮した素性を加えた場合の要約結果
Table 5　Summarization result with CP feature and features which take into account consecutiveness of important sentences and redundancy.

Trn.	Condition	κ	F	Rouge-4
Manual	従来の素性	0.388	0.548	0.696
	従来の素性+CP	0.382	0.544	0.692
	＋ ①	0.384	0.545	0.693
	＋ ②	0.394	0.552	0.706
	＋ ① ＋ ②	0.401	0.558	0.711
	＋ ① ＋ ② ＋ ③	0.404	0.560	0.711
ASR	従来の素性	0.375	0.538	0.680
	従来の素性+CP	0.381	0.543	0.689
	＋ ①	0.384	0.545	0.691
	＋ ②	0.381	0.542	0.694
	＋ ① ＋ ②	0.395	0.553	0.702
	＋ ① ＋ ② ＋ ③	0.391	0.550	0.699
Human		0.469	0.597	0.695

※ 従来の素性は 3.2.1 節中の全素性．①，②，③ はそれぞれ動的素性，差分素性，冗長素性．

また，CP に基づいて抽出された文数は，人手による書き起こしを使用した場合には平均 161.6 文（17%），音声認識結果を使用した場合には平均 129.5 文（13%）であった．κ 値が小さいのは抽出文数が少ないためである．

4.4.1.3　CP 抽出結果に基づく素性の自動要約への適用結果

CP 素性を 3.2.1 項の素性と組み合わせた場合の結果を**表 5** に示す．CP 素性を用いることで，人手による書き起こしを用いる場合には逆に評価値が下がる結果となったが，音声認識結果を用いる場合には評価値の向上が得られた．人手による書き起こしを使用した場合に CP 素性の効果がなかったのは，人手による書き起こしには単語の認識誤りがないために他の言語情報による文抽出の精度が高く，新たな言語情報により改善できる余地が少なかったためであると考えられる．逆に，音声認識結果を使用した場合に CP 素性の効果があったのは，新たな言語情報（単語/品詞列のパターン）によって改善できる余地があったためであると考えられる．音声認識結果を使用する場合に効果が見られたことから，CP 素性は音声認識誤りに頑健な素性であるといえる．音声認識誤りに対して CP 素性がうまく働いた例としては，「した時、する時 には、人間は何 か を決めてやらなきゃいけないと」を「辿りとする時 には 復元何 か と決めてやらなきゃいけないと」と誤認識した場合でも，「に は *

か」という CP が抽出されることで重要文として抽出できた場合などがあった．

4.4.2　重要文の連続性を考慮した要約の評価

表 5 に，3.2.1 項の素性と CP 素性に加えて，4.2.2 項に示した連続性を考慮する素性を加えた場合の結果を示す．4.2 節では，連続性を考慮する素性として動的素性と差分素性の 2 つを提案したが，動的素性または差分素性のみを単独で加えた場合には，評価値に与える影響は小さく，それぞれ単独ではあまり有効な素性ではないといえる．しかし，両方の素性を同時に使用すると，人間による書き起こしを使用した場合には κ 値で 0.019，F 値で 0.014，*Rouge*-4 で 0.019，音声認識結果を使用した場合には κ 値で 0.014，F 値で 0.010，*Rouge*-4 で 0.013 の向上が見られ，大幅な評価値の向上が見られた．両方の素性を同時に使用することで効果が得られたのは，動的素性と差分素性の性質によって，両者が相補的に働くからであると考えられる．動的素性は，基本的に，文をまとめて抽出した場合にスコアが高くなるように働き，一方，差分素性は，直前の文と素性の値が似ている場合にスコアが高くなるように働く．したがって，直前の文と素性の値が似ている文が連続して存在する場合を考えると，差分素性を考慮することで，塊としてこれらの文のスコアが大きくなり，ここで，さらに動的素性を考慮すると，単独ではスコアがより大きい他の文を抽出するよりも，この連続した文の塊を抽出した方が全体としてスコアが大きくなる可能性が高くなる．このように，差分素性が連続した文の塊を強調し，動的素性がこれをまとめあげる働きをすることで，両者が相補的に働くものと考えられる．

4.4.3　冗長性を考慮した要約の評価

表 5 に，前項までの素性に加えて，4.3.1 項に示す冗長性素性を加えた場合の結果を示す．冗長性素性を使用することで，人間による書き起こしを使用した場合には若干の向上が見られたが，音声認識を使用した場合には向上が見られなかった．冗長性素性によって結果が向上しなかった主な原因として，冗長性を文を単位として計算しているために，塊として冗長な部分を排除できていない可能性があげられる．冗長性を文の塊に対して考慮するためには，素性の設計段階から修正する必要がある．自動トピック分割の結果を利用するなど，冗長性を文の塊に対して扱う方法に関しては，今後さらなる検討が必要である．

4.4.4　書き起こしによる差異の比較

提案した 3 手法による素性をすべて使用した場合に，人手による書き起こしを使用した場合の結果と音声認識結果を使用した場合の結果の差分は，それぞれの評価尺度において，表 5 より $\Delta\kappa = 0.013$，$\Delta F = 0.010$，Δ*Rouge*-4 $= 0.012$ であった．これらの値は，表 1 に示す Accuracy の平均が 49.1%，Correct の平均が 55.8%であったことを考えると，大変小さな値であるといえる．したがって，本論文で提案した feature-based の改善手法は，音声認識誤りに対して十分頑健であるといえる．表 2 に示すとおり，MMR 法および，従来の素性のみを用いた feature-based においても，書き起こし間の差異は少なく，音声認識誤りに対して頑健であるといえる．今回の実験における WER は約 50%であるため，これらの結果は，WER が 50%程度と高い場合においても，単語誤りは音声要約の精度にあまり影響を与えないことを示した文献 4) と同様の結論を与えるものである．

4.4.5　人間の要約との比較

提案した 3 手法による素性をすべて使用した場合の要約結果と，人間による要約結果を比較する．κ 値および F 値においては，人間による書き起こしで $\kappa = 0.404$，$F = 0.560$，音声認識結果で $\kappa = 0.391$，$F = 0.550$ であるのに対し，人間による要約は $\kappa = 0.469$，$F = 0.597$ であり，依然及ばない結果となった．しかし，Rouge-4 では，機械による要約が，人間による書き起こしでは 0.711，音声認識結果では 0.699 であり，人間の要約である 0.695 を上回っている．Rouge は内容の保存に関する尺度であるため，機械による要約は内容に関しては保存できているといえる．

5. 結　　論

本論文では，重要文抽出に基づいた講義音声の自動要約に関して述べた．音声要約で広く使用される MMR と feature-based による要約手法を比較した結果，feature-based による要約手法の方が優れた性能を示した．さらに，feature-based による要約の精度を高めるため，3 つのアプローチを試みた．1 つ目は，高精度に各文の重要度を表す素性の抽出を試みるアプローチで，重要文の手がかり表現（CP）を自動抽出し，自動要約の素性とする手法であった．CP 抽出結果に基づく素性を用いることで，音声認識結果を使用した場合には要約の改善を得ることができた．2 つ目は，人間による要約は重要文が連続しやすいという観測に基づいて，この連続性をとらえる素性を使用することで要約を改善しようとするアプローチであった．動的素性と差分素性が互いに相補的に働くことで，重要文の連続性をとらえることができ，要約の改善を得ることができた．3 つ目は，冗長性を排除する枠組みの導入であった．正解書き起こしを使用した場合には若干の向上が見られたが，音声認識結果を使用した場合にはあまりうまく働かなかった．冗長性の排除に関しては，まだ改善の余地がある．

今後の課題としては，より効率的に冗長性を排除できる素性・枠組みの考案と，より高度なコンテキスト情報を素性として組み込む方法の考案があげられる．

参 考 文 献

1) Glass, J., Hazen, T.J., Hetherington, L. and Wang, C.: Analysis and processing of lecture audio data; Preliminary investigations, *Proc. HLT-NAACL 2004*, pp.9–12 (2004).
2) Lamel, L., Adda, G., Bilinski, E. and Gauvain, J.L.: Transcribing Lectures and Seminars, *Proc. Interspeech*, pp.4–8 (2005).
3) 富樫慎吾，山口　優，北岡教英，中川聖一：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).
4) Zhu, X. and Penn, G.: Summarization of Spontaneous Conversations, *Interspeech*, pp.1531–1534 (Sep. 2006).
5) Chen, Y., Chiu, H., Wang, H. and Chen, B.: A Unified Probabilistic Generative Framework for Extractive Spoken Document Summarization, *Interspeech*, pp.2805–2808 (2007).
6) Daniel, R. and Martins, D.: Extractive Summarization of Broadcast News: Comparing Strategies for European Portuguese, *TSD*, Vol.4629, pp.115–122, Springer (2007).
7) 中川聖一，富樫慎吾，山口　優，藤井康寿，北岡教英：講義音声ドキュメントのコンテンツ化と視聴システム，電子情報通信学会誌，Vol.91-D, No.2, pp.238–249 (2008).
8) 小林　聡，山口　優，中川聖一：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).
9) Xie, S. and Liu, Y.: Using Corpus and Knowledge-based Similarity Measure in Maximum Marginal Relevance for Meeting Summarization, *ICASSP*, pp.4985–4988 (2008).
10) Liu, Y. and Xie, S.: Impact of Automatic Sentence Segmentation on Meeting Summarization, *ICASSP*, pp.5009–5012 (2008).
11) 南條浩輝，北出　祐，河原達也：談話標識の統計的選択に基づいた CSJ の講演からの重要文抽出，電子情報通信学会技術研究報告 NLC，言語理解とコミュニケーション，Vol.103, No.517, pp.73–78 (2003).
12) 岩野公司，広畑　誠，新中庸介，古井貞熙：重要文抽出による音声自動要約手法とその客観評価法についての検討，電子情報通信学会技術研究報告 SP，音声，Vol.105, No.132, pp.1–6 (2005).
13) Kikuchi, T., Furui, S. and Hori, C.: Automatic Speech Summarization Based on Sentence Extraction and Compaction, *ICASSP*, pp.384–387 (2003).
14) 土屋雅稔，小暮　悟，西崎博光，太田健吾，山本一公，中川聖一：日本語講義音声コンテンツコーパスの作成と分析，情報処理学会論文誌，Vol.50, No.2, pp.448–450 (2009).
15) Murray, G., Renal, S. and Carletta, J.: Extractive Summarization of Meeting Recording, *Interspeech*, pp.593–596 (2005).
16) Togashi, S., Yamaguchi, M. and Nakagawa, S.: Summarization of Spoken Lectures Based on Linguistic Surface and Prosodic Information, *IEEE/ACL Workshop on Spoken Language Technology*, pp.34–37 (2006).
17) 北岡教英，高橋伸寿，中川聖一：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).
18) Furui, S., Maekawa, K. and Isahara, H.: A Japanese National Project on Spontaneous Speech Corpus and Processing Technology, *Proc. ASR2000*, pp.244–248 (2000).
19) 富樫慎吾，中川聖一：講義音声ドキュメントのコンテンツ化とブラウジングシステムの改良，*Proc. 2nd Spoken Document Processing Workshop*, pp.155–160 (2008).
20) Lin, C. and Hovy, E.: Automatic Evaluation of Summaries Using N-gram Co-Occurence Statistics, *The Human Language Technology Conference*, pp.71–78 (2003).
21) Nenkova, A.: Summarization Evaluation for Text and Speech: Issues and Approaches, *Interspeech*, pp.1527–1531 (2006).
22) Fleiss, J.L.: Measuring Nominal Scale Agreement Among Many Rater, *Psychological Bulletin*, Vol.76, pp.378–382 (1971).
23) Carbonell, J. and Goldstein, J.: The Use of MMR, Diversity-Based Reranking for Reordering Documents and producing summaries, *Proc. ACM SIGIR*, pp.335–336 (1998).
24) 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).
25) Entropicspeech Technology: *ESPS Manual Pages* (1998). http://www.ee.uwa.edu.au/∼roberto/research/speech/local/entropic/ESPSDoc/manpages/indexes/
26) Hirao, T., Isozaki, H., Maeda, E. and Matsumoto, Y.: Extracting Important Sentences with Support Vector Machines, *Proc. COLING*, pp.342–348 (2002).
27) Vapnik, V.N.: *The Nature of Statistical Learning Theory*, Springer, New York, NY, USA (1995).
28) Joachims, T.: Training Linear SVMs in Linear Time, *Proc. ACM Conference on Knowledge Discovery and Data Mining* (*KDD*), pp.217–226 (2006).
29) Lin, S.-H., Chen, Y.-T., Wang, H.-M. and Chen, B.: A Comparative Study of Probabilistic Ranking Models for Spoken Document Summarization, *ICASSP*, pp.5025–5028 (2008).
30) Lafferty, J. and McCallum, F.P.A.: Conditional Random Fields: Probabilistic Models for Segmenting and Labeling Sequence Data, *Proc. 18th International Conference on Machine Learning* (2001).

(平成 21 年 5 月 14 日受付)
(平成 21 年 12 月 17 日採録)

藤井　康寿（学生会員）

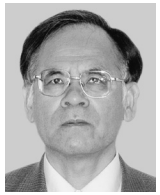
平成 19 年豊橋技術科学大学情報工学系卒業，平成 21 年同大学大学院修士課程情報工学専攻修了．現在，同大学院博士後期課程電子・情報工学専攻在学中．主として音声認識および音声要約に関する研究に従事．日本音響学会，電子情報通信学会，IEEE 各会員．

山本　一公（正会員）

平成 7 年豊橋技術科学大学情報工学系卒業，平成 9 年同大学大学院修士課程情報工学専攻修了．平成 12 年同大学院博士後期課程電子・情報工学専攻修了．同年信州大学工学部助手．平成 19 年豊橋技術科学大学情報工学系助教．博士（工学）．主として音声認識に関する研究に従事．日本音響学会，電子情報通信学会各会員．

北岡　教英（正会員）

平成 4 年京都大学工学部情報工学科卒業．平成 6 年同大学大学院修士課程修了．同年（株）デンソー入社．平成 9〜12 年豊橋技術科学大学大学院博士後期課程在学．平成 13 年豊橋技術科学大学情報工学系助手．平成 15 年同講師．平成 18 年名古屋大学大学院情報科学研究科助教授．平成 19 年同准教授．博士（工学）．平成 21 年 Nanyang Technologycal University（Singapore）Visiting Assosiate Professor．主として音声認識，音声対話，音声インタフェースに関する研究に従事．ISCA，電子情報通信学会，日本音響学会，人工知能学会各会員．



中川　聖一（フェロー）

昭和 51 年京都大学大学院博士課程修了．同年同大学情報工学科助手．昭和 55 年豊橋技術科学大学情報工学系講師．平成 2 年同大学教授．カーネギーメロン大学客員研究員（昭和 61〜62 年）．音声情報処理，自然言語処理，人工知能の研究に従事．工学博士．昭和 52 年電子通信学会論文賞，昭和 63 年度 IETE 最優秀論文賞，平成 13 年電子情報通信学会論文賞受賞，電子情報通信学会フェロー．情報処理学会フェロー．著書『確率モデルによる音声認識』（電子情報通信学会編），『音声・聴覚と神経回路網モデル』（共著，オーム社），『情報理論の基礎と応用』（近代科学社），『パターン情報処理』（丸善）等．