2008－SLP－72 (16)
2008／7／19

重要文の連続性を考慮した講義音声の自動要約

藤井 康寿　　山本 一公　　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
E-mail： {fujii,kyama,nakagawa}@slp.ics.tut.ac.jp

概要

本稿では，SVMを用いた重要文の連続性を考慮した重要文抽出手法について報告する．一般的に，重要文抽出による音声自動要約は，要約全体の冗長性に関しては考慮していても，文間の関係性は考慮していない．しかし，抽出された文の間には関係性が存在するはずであり，実際に，重要文の連続性として文間の関係が観測される．本稿では，このような連続性を考慮するような素性を feature-based の要約に取り入れることで，重要文抽出による要約を改善する方法を提案する．講義音声コンテンツコーパス CJLC を用いた実験の結果，重要文の連続性を考慮することで重要文抽出による要約を改善できることを示す．また，MMR 法に基づいた素性を導入することで，要約全体の冗長性を取り除く方法についても提案する．

キーワード 講義音声，音声自動要約，重要文抽出，重要文の連続性

Improvement of Class Lecture Summarization by Taking into Account Consecutiveness of Important Sentences

Yasuhisa FUJIII Kazumasa YAMAMOTO Seiichi NAKAGAWA
Department of Information and Computer Sciences, Toyohashi University of Technology
1-1, Hibarigaoka, Tempaku-cho, Toyohashi 441-8580, Japan
E-mail： {fujii,kyama,nakagawa}@slp.ics.tut.ac.jp

Abstract

This paper presents a novel sentence extraction framework that takes into account the consecutiveness of important sentences using a Support Vector Machine (SVM). Generally, most extractive summarizers do not take context information into account, but do take into account the redundancy over the entire summarization. However, there must exist relationships among the extracted sentences. Actually, we found these relationships as consecutiveness among the sentences. We deal with this consecutiveness by using new two features for a feature-based summarizer. Experimental results on a Corpus of Japanese classroom Lecture Contents (CJLC) showed that our proposed method outperformed traditional methods, which did not take context information into account. We also present a way to ensure based on MMR that no redundant summarization occurs.

key words Classroom lecture speech, Automatic speech summarization, Sentence extraction, Consecutiveness of important sentences

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, 4, 5, 6, 7, 8]．

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が，本稿では抽出型の要約を対象とする．抽出型の要約のうち，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要約と呼ぶ．

重要文抽出による音声要約手法は，世界中で広く研究されている．Zhu ら [4] は，音声自動要約における言い淀みの役割と各素性に対する WER の影響を調査した．Chen ら [5] は文のランキングのために文からのドキュメントの生成確率と文の事前確率を統合する方法を提案し，確率モデルによって要約を作成する手法を示した．Ricardo ら [6] は，文抽出要約において代表的な手法である feature-based による要約手法，Latent Semantic Analysis (LSA) による要約手法，Maximal Marginal Relevance (MMR) による要約手法を比較した．このように，テキスト要約のベースラインとなっている MMR は，音声要約でもよく用いられている [9, 10]．

一般的に，重要文抽出に基づく要約は，文献 [4, 5, 6, 7] に示される様に，抽出された文がそれぞれ独立である

と仮定しており，文間の関係性（コンテキスト）を考慮していない．しかし，文どうしは互いに関係しあっているはずであり，より良い要約の生成のためにコンテキスト情報は当然利用されるべきである．文献 [11] において，Kolluru らは人間による重要文抽出結果が連続し易いことに着目し，重要と判定された文の回りも同時に抽出する要約手法を試した．しかし，人間による主観評価の結果，単純に回りの文をまとめて抽出しただけだったため，一貫性に関しては向上したが，コンテキストを考慮しない方法に比べて良い結果を得ることができなかった．Maskey ら [12] は，音声の要約において，HMM の隠れ状態で文の抽出／非抽出を表現し，直前の抽出に関する決定を現在の決定に反映することができる要約手法を提案した．しかし，Maskey らの方法は言語情報を一切使用しておらず，要約に有効な多くの素性を考慮できていない．

本稿では，人間の要約における重要文の連続性に着目し，重要文の連続性を考慮した要約手法を提案する．重要文の連続性をモデル化するために，動的素性と差分素性を新たに素性として使用する．動的素性とは，直前の文の抽出結果に基づく素性であり，差分素性とは，現在の文と直前の文の素性の値の差分である．ドキュメント中の各文は，動的素性，差分素性を含めた文の重要度を決定するための素性に基づいて，Support Vector Machine（SVM）によって分類される．本稿では，上述の枠組みの中で MMR 法に基づいて冗長性を取り除くための方法も提案する．実験結果により，コンテキスト情報を考慮した提案法が，コンテキスト情報を考慮しない方法を上回ることを示す．

2 重要文の連続性

2.1 音声試料

本稿では，日本語講義音声コンテンツコーパス CJLC [13] 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 1000 文からなる．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．要約実験は，人手による書き起こしと音声認識結果の両方に対して行う．人手による書き起こしは，CJLC に付随しているため，これを利用した．音声認識結果には，SPOJUS [14] 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 [15] から学習した．本稿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 [16].

2.2 要約の正解

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% の文を抽出する様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を正解文とした．すなわち，3 人以上の被験者が正解と判定した文の集合を要約の正解とした．本稿ではこれを **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 [17].

表 2 重要文の連続性

重要文数	孤立重要文数	平均	最大
268.0	80.6	1.83	7.0

要約の目標値は，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 人の被験者の内 3 人以上が重要と判定した文集合（**man3/5**）の間の一致度（κ 値および F 値．次節にて説明）の全被験者間の平均とした．値を表 1 に示す．被験者間の一致度よりも，ある被験者とその被験者を除いた man3/5 の一致度の方が高い値を示す [17] [*1].

2.3 重要文の連続性

表 2 は，**man3/6** における重要文の連続性を示す．**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．この観測より，約 70% の重要文は連続して出現していることがわかる [*2]．重要文の平均連続長は 1.83 であり（各文を独立に 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わかる．もし，この表面的な重要文の連続性の観測を上手く捉えることができれば，より高度な要約を作成することができる可能性がある．

3 Baseline 手法

3.1 Maximal Marginal Relevance による要約

Maximal Marginal Relevance（MMR）[18] 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る．元々はテキスト要約を対象とした手法であるが，文献 [19] 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，本稿では文献 [19] で定義されるものを使用する．MMR による文 S_i のスコア $S_C^{MMR}(i)$ は以下の式で与えら

[*1] 任意の被験者の要約文集合間の κ 値は平均 0.387 であった．

[*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（各文を独立に 1/3 抽出した場合の期待値は 1.50）．

表 1 音声試料の諸元

Duration	No.of Snt.	Accuracy [%]	Correct [%]	要約の目標値 (対 man3/5)		
				κ	F	ROUGE-4
67'55"	973.8	49.1	55.8	0.469	0.597	0.695

れる.

$$S_C{}^{MMR}(i) = \lambda(Sim(\boldsymbol{S}_i, \boldsymbol{D})) - (1-\lambda)(Sim(\boldsymbol{S}_i, \boldsymbol{S_{rk}})) \ , \quad (1)$$

ここで，$\boldsymbol{D}$ はドキュメントに含まれる全ての文の平均ベクトルであり，$\boldsymbol{S}_{rk}$ はこれまでに抽出した文集合の平均ベクトルである．Sim は 2 つのベクトル間のコサイン距離であり，2 つのベクトル間の類似度を示す．式 (1) の第一項は文とドキュメントの関連度を表し，第二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．$\boldsymbol{S}_i$ は文 i に含まれる単語からなるベクトルであり，本稿においては以下のように定義する．

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

$$tf_{i,w} = f_w \cdot \log(\frac{f_{\hat{w}}}{f_w}) \ , \quad (3)$$

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．

実験では，式 (1) 中の λ について，$0.0 \sim 1.0$ 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．

3.2 Feature-based による要約

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 節において我々が従来から使用している素性 [17] および素性の処理に関して説明し，3.2.5 節において使用する識別器について説明する．

3.2.1 素性

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 [17, 7].

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 [19]．言語素性には，単語（正確には形態素）に基づくものが多くあるが，本研究では，形態素解析器には *ChaSen* [20] を使用した．本稿で使用する言語素性は以下の通りである．

Repeated words: 頻出単語を含む数を素性とする．頻出単語はある一定の閾値以上の出現頻度を持つ語である．閾値は，ドキュメント中の単語を頻度の多いものから順に使用し，それらの単語を文中に 2 語以上含む文を抽出したときに，丁度設定要約率に達する様に設定する．

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．よって，該当の文中でスライドに含まれる単語が発話された回数を素性として使用する．

Term Frequency (TF): フィラーや不要語を除いて，式 (3) の値を TF として使用する．

CP-based: 先行研究 [7] に示される方法で各文に CP（Cue Phrase）ラベルを付与し，ラベルが付与された個数を素性とする．

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することが可能である [17]．本稿で使用する韻律情報は以下の通りである．

Duration: 各文の発話時間長を素性として使用する．

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．本稿では，パワーおよび F0 の値は ESPS [21] 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．本稿にける話速とは，単位時間あたりのモーラ数であり，文 i に対する話速 $ROS(i)$ は以下の式で計算する：

$$ROS(i) = \frac{\sum_{w \in S_i} mora(w)}{duration(i)}, \quad (4)$$

ここで，$mora(w)$ は単語 w のモーラ数であり，$duration(i)$ は文 i の発話時間長である．

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 [17].

3.2.2 素性の正規化

3.2.1 節に示す素性は，講義や話者によってばらつきがあるため，講義毎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．正規化は以下の式によって行う．

$$f_j{}^{norm}(S_i) = \frac{f_j(S_i) - mean(f_j)}{std(f_j)}, \quad (5)$$

ここで，$f_j(S_i)$ は文 S_i の素性 j の元々の値であり，$mean(f_j)$ および $std(f_j)$ は対象の講義における素性 j の平均および標準偏差である．

3.2.3 素性の拡張

素性の表現力を高めるために素性の拡張を行う．本稿では，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．

3.2.4 素性の量子化

文献 [22] と同様に，本稿では素性の値を div 個の値で量子化し，全ての素性を div 個の2値変数で表現する．あるドキュメント中における素性 j の最大値が max_j，最小値が min_j であるとき，文 i に対する素性 j の値 $f_j(i)$ は以下の式で量子化される．

$$f_j{}^{disc} = rounddown\left(\frac{f_j{}^{norm} - min_j}{max_j - min_j} \cdot div\right), \quad (6)$$

ここで，$rounddown$ は小数点以下切り捨て関数である．これにより，f_j^{norm} は div に量子化されるので，その量子化された値を div bit の値の各桁に対応させ，対応の bit のみを 1，それ以外を 0 とする．例えば，$f_j{}^{disc} = 0$ は "10000"，$f_j{}^{disc} = 1$ は "01000"，$f_j{}^{disc} = 4$ は "00001" と変換される．

3.2.5 識別器

文 i のスコアは以下の式で計算する．

$$Score(S_i) = \boldsymbol{w}\boldsymbol{x} + b\,, \quad (7)$$

ここで，$\boldsymbol{x}$ は 3.2.1 節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 [23] の線形カーネルによって学習する．SVM の学習には，svm^{perf} [24] を用いた．

実験では，式 (7) の $\boldsymbol{w}$，b の学習は．話者に対してオープンとなる様に行った．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 提案法

4.1 Feature-based による重要文の連続性を考慮した要約

2.3 節で示した通り，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．この重要文の連続性を扱うために，本稿において動的素性および差分素性と呼ぶ 2 つの素性を新たに提案する．

4.1.1 動的素性

重要文に連続して抽出され易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．よって，本稿では動的素性として，直前の文の抽出結果を使用する．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．

$$dynamic(i) = \begin{cases} 10 & if \quad S_{i-1} \text{ is extracted.} \\ 01 & otherwise. \end{cases} \quad (8)$$

4.1.2 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話しているからである．従って，重要文のまとまりの中で，近接する文どうしは似たような素性の値を持っているはずである．反対に重要文と非重要文の境界でとなり合う文どうしは異なった素性の値を持っているはずである．この予測に基づいて，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る．文 i からの素性 j に対する差分素性 $diff_{i,j}$ は以下の様に計算される．

$$diff_{i,j} = f_j(S_i) - f_j(S_{i-1}). \quad (9)$$

4.1.3 最適系列の同定

連続性を考慮した要約では，現在の決定は直前の文の決定に依存しているため，文毎に独立にスコアを計算することができない．しかし，今回の仮定では現在の文の抽出は直前の文の抽出にのみ依存しているので，下記の動的計画法を解くことで一意に全体のスコアを最大化する系列を同定可能である．

$$g_0(i,j) = \max \begin{cases} g_0(i-1,j) \\ g_1(i-1,j) \end{cases} \quad (10)$$

$$g_1(i,j) = \max \begin{cases} g_0(i-1,j-1) + score(i|0) \\ g_1(i-1,j-1) + score(i|1), \end{cases} \quad (11)$$

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 番目の文までの抽出において 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 (7) 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．

4.2 Feature-based による冗長性を考慮した要約

4.1 節において，feature-based による要約において重要文の連続性を考慮するための方法を提案したが，4.1 節の方法は冗長性を考慮していない．本節では，feature-based による要約において冗長性を排除する枠組みを提案する．冗長性を排除する為に，MMR における冗長性排除の基準を素性として feature-based による要約に適用する．

4.2.1 冗長性を扱う素性

式 (1) の第 2 項は，既に抽出した文集合と抽出対象の文との類似度を示し，冗長性を示している．本稿ではこの冗長性を素性として使用する．人間による要約は冗長性が無く，各文の間に関連性はあったとしても，それぞれが保持している情報は独立であると仮定すると，冗長性が高いことによって選ばれなかった文は人

間による重要文集合と類似度が高くなり，そうでない場合は類似度が低くなるといえる．この仮定に基づいて，ドキュメントとそれに対する重要文集合 $\boldsymbol{imp}$ が与えられたときの，文 i に対する冗長性素性 $rdun(i)$ を下記の様に計算する．

$$rdun(i) = Sim(S_i, Imp), \tag{12}$$

$$Imp = \begin{cases} \frac{\sum_{S \in \boldsymbol{imp} - S_i} S}{|\boldsymbol{imp}|} & if \quad S_i \in \boldsymbol{imp} \\ \frac{\sum_{S \in \boldsymbol{imp}} S}{|\boldsymbol{imp}|} & otherwise, \end{cases} \tag{13}$$

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．

4.2.2 探索

4.2.1 節における素性を使用することで，提案法は要約全体に依存関係を定義していることになるので，全ての仮説から最適な系列を同定することはほぼ不可能である．系列の探索は，式 (10) 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によって行う．また，式 (13) における $\boldsymbol{imp}$ は正解文集合であり使用することができないので，$\boldsymbol{imp}$ は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る．実験では，ビーム幅 W を 30 とした．

5 評価実験

5.1 評価尺度

評価尺度には，κ 統計量（κ 値） [25]，F 値，および *Rouge*-N [26] を用いる．それぞれ次のように定義される．

- **κ 値** [25]
  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：

$$\kappa = \frac{P(A) - P(E)}{1 - P(E)} \tag{14}$$

$$P(A) = \quad A \text{ と } B \text{ の一致率} \tag{15}$$

$$P(E) = A \text{ と } B \text{ の偶然の一致率} \tag{16}$$

- **F 値**
  F 値は Precision（精度）と Recall（再現率）の調和平均として定義される：

$$F\text{-}measure = \frac{2 \times Precision \times Recall}{Precision + Recall}, \tag{17}$$

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

  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** [26]
  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計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，どれか一箇所が抽出されれば高い評価値となる利点がある．

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

  本稿では N=4 とし，*Rouge*-4 を使用する．

5.2 要約結果

表 3 に各条件における要約結果を示す．

まず，feature-based による要約と MMR による要約を比較する．表より，feature-based による要約が MMR よりも明らかに良い性能を示すことがわかる．この結果は，文献 [19] や [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，これは，本稿における方法の方がより高度な feature-based 手法を採用していること，講義はスライドに基づいて行われているため，会議や放送ニュースに比べて冗長性が少ないことによるものと考えられる．

次に，連続性を考慮する素性について考察する．連続性を考慮する素性として動的素性または差分素性のみを使用した場合には，評価値に与える影響は小さく，あまり有効な素性ではない．しかし，動的素性と差分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も音声認識結果を使用した場合についても大幅な評価値の向上が見られた．この結果は，差分素性が重要文の塊を強調し，動的素性がそれらをまとめ上げる働きをすることによって，両者が相補的に働いているものと考えられる．

次に，冗長性素性を加えた結果について考察する．冗長性素性を従来の素性に単独で加えた場合には，音声認識結果を使用した場合の Rouge 値のみ向上が見られた．また，冗長性素性を，連続性を考慮する素性と同時に使用すると，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．以上の結果より，冗長性素性はあまりうまく働いていないといえる．これは，探索がうまく出来ていないか，もしくは，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ことによると考えられる．

最後に，それぞれの書き起こしにおいて一番良い値だった結果と人間の要約について比較する．κ 値および F 値においては，正解書き起こしで $\kappa = 0.404$, $F = 0.560$，音声認識結果で $\kappa = 0.395$，$F = 0.553$ 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 においては機械による要約の方が勝っている．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．

各ケースにおいて，正解書き起こしを使用した場合と音声認識結果を使用した場合の結果は非常に近いも

表 3 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

Trn.	Condition	κ	F	Rouge-4
Manual	MMR	0.342	0.511	0.625
	従来の素性	0.382	0.544	0.692
	+ ①	0.384	0.545	0.693
	+ ②	0.394	0.552	0.706
	+ ③	0.382	0.544	0.693
	+ ① + ②	0.401	0.558	0.711
	+ ① + ② + ③	0.404	0.560	0.711
ASR	MMR	0.338	0.508	0.627
	従来の素性	0.381	0.543	0.689
	+ ①	0.384	0.545	0.691
	+ ②	0.381	0.542	0.694
	+ ③	0.380	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．

のとなっているので，我々の要約手法は音声認識誤りに対して頑健であるといえる．

提案した素性（動的素性，差分素性，冗長素性）が，人間の主観評価に対しても有効かどうかを調査するために，音声認識結果を使用して，従来素性のみを使用した場合の要約音声と，従来素性に加えて提案素性を使用した場合の要約音声を作成し，被験者 10 名による聴取実験を行った．質問内容は，“講義の内容がつかみやすかったかどうか（講義の要点が保存されているか）”，“自然な音声に聞こえたか（文のつながり・流れ等が自然であるかどうか）”の 2 点であった．実験の結果，両方の観点において，両者間に有意な差は見られなかった．人間の主観評価の上で品質の高い要約を作成することは今後の課題である．

6 おわりに

本稿では，feature-based による要約において，動的素性および差分素性を使用することで重要文の連続性を考慮する重要文抽出手法に関して述べた．動的素性とは，直前の文の抽出結果に基づく素性であり，差分素性は現在の文と直前の文の間の素性値の差分である．動的素性と差分素性は相補的であり，同時に使用することで連続性を考慮した素性を使用しない場合に比べて良い要約を生成することができた．また，feature-based による要約において冗長性を排除する枠組みも提案した．冗長性を排除する素性を使用することで，人手による書き起こしを使用した場合には若干の改善を得ることができた．しかし，冗長性の排除に関してはまだ改善の余地がある．提案法は，各評価尺度の上では有効であり，評価値の向上が得られたが，人間の主観評価の上では有意な改善が得られなかった．

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上げられる．また，人間の主観評価の上で品質の高い要約を作成することも今後の課題である．

参考文献

[1] J. Glass, T. J. Hazen, L. Hetherington, and C. Wang. Analysis and processing of lecture audio data; preliminary investigations. In *Proceedings of the HLT-NAACL 2004*, pp. 9–12, 2004.

[2] L. Lamel, G. Adda, E. Bilinski, and J. L. Gauvain. Transcribing lectures and seminars. *Interspeech*, pp. 4–8, September 2005.

[3] 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.

[4] X. Zhu and G. Penn. Summarization of spontaneous conversations. *Interspeech*, pp. 1531–1534, 2006. 9.

[5] Y. Chen, H. Chiu, H. Wang, and B. Chen. A unified probabilistic generative framework for extractive spoken document summarization. *Interspeech*, pp. 2805–2808, August 2007.

[6] R. Daniel and D. Martins. Extractive summarization of broadcast news: Comparing strategies for european portuguese. In *TSD*, Vol. 4629, pp. 115–122. Springer, September 2007.

[7] 藤井, 北岡, 中川. 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. 情報処理学会研究報告, SLP-66-15, 2007. 5.

[8] 中川, 富樫, 山口, 藤井, 北岡. 講義音声ドキュメントのコンテンツ化と視聴システム. 電子情報通信学会論文誌, Vol. 91-D, No. 2, pp. 238–249, 2 2008.

[9] S. Xie and Y. Liu. Using corpus and knowledge-based similarity measure in maximum marginal relevance for meeting summarization. In *ICASSP*, pp. 4985–4988, 2008.

[10] Y. Liu and S. Xie. Impact of automatic sentence segmentation on meeting summarization. In *ICASSP*, pp. 5009–5012, 2008.

[11] B. Kolluru, H. Christensen, Y. Gotoh, and S. Renals. Exploring the style-technique interaction in extractive summarization of broadcast news. *IEEE ASRU Workshop*, 2003.

[12] S. Maskey and J. Hirschberg. Summarizing speech without text using hidden markov models. In *Proceedings of the HLT-NAACL 2006*, pp. 89–92. ACL, June 2006.

[13] 小暮, 西崎, 土屋, 富樫, 山本, 中川. 日本語講義音声コンテンツコーパスの構築と講義音声認識手法の検討. Proc. 第 2 回音声ドキュメント処理ワークショップ, pp. 7–14, 2008.

[14] 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.

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

[16] 富樫, 中川. 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. 第 2 回音声ドキュメント処理ワークショップ, pp. 155–160, 2008.

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

[18] J. Carbonell, Y. Geng, and J. Goldstein. Automated query-relevant summarization and diversity-based reranking. *AI and Digital Libraries*, pp. 9–14, 1997.

[19] G. Murray, S. Renal, and J. Carletta. Extractive summarization of meeting recording. *Interspeech*, pp. 593–596, 2005.

[20] 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.

[21] entropic speech technology. *ESPS Manual Pages: http://www.ee.uwa.edu.au/~roberto/research/speech/local/entropic/ESPSDoc/manpages/indexes/*, 1998.

[22] T. Hirao, H. Isozaki, E. Maeda, and Y. Matsumoto. Extracting important sentences with support vector machines. *Proceedings of COLING*, pp. 342–348, 2002.

[23] 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.

[24] T. Joachims. Training linear SVMs in linear time. In *Proceedings of the ACM Conference on Knowledge Discovery and Data Mining (KDD)*, pp. 217 – 226, 2006.

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

[26] C.-Y. Lin. Rouge: a package for automatic evaluation of summaries. In *Proceedings of the Workshop on Text Summarization Branches Out (WAS 2004)*, pp. 74–81, July 2004.