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

2003－NL－155 (13)
2003－SLP－46 (2)
2003／5／27

談話標識と話題語に基づく統計的尺度による講演からの重要文抽出

北出 祐　南條 浩輝　河原 達也　奥乃 博

京都大学 情報学研究科 知能情報学専攻
〒 606-8501 京都市 左京区 吉田二本松町
e-mail: kitade@ar.media.kyoto-u.ac.jp

あらまし　講演（学会講演）のディジタルアーカイブ化を目的として，書き起こし（音声認識結果）から自動的に重要文を抽出するために，学会講演特有の話題構造を利用した談話標識に基づく手法を提案する．ポーズ情報および言語的情報をもとに話し言葉におけるセクション境界候補を検出し，セクション冒頭の文に頻出する談話標識を求めた上で，これに基づく統計的な重要度尺度を定義する．さらに話題語（キーワード）の統計量に基づく重要度尺度と統合することも検討した．これらの重要度尺度で CSJ の 14 件の学会講演を対象に重要文抽出精度の評価を行い，(1) 談話標識に基づく手法が有効であること，(2) 話題語に基づく手法と統合することで相乗効果が得られること，を確認した．

キーワード　講演, 重要度尺度, 談話標識, 話題語, 重要文抽出

Automatic Extraction of Important Sentences from Lecture Transcription using Statistical Measure based on Discourse Markers and Topic Words

Tasuku Kitade　Hiroaki Nanjo　Tatsuya Kawahara　Hiroshi G. Okuno

School of Informatics, Kyoto University, Kyoto 606-8501, Japan
e-mail: kitade@ar.media.kyoto-u.ac.jp

Abstract

For efficient access to speech media, secondary information is required. We explore automatic extraction of important sentences from lecture presentations. We segment a lecture into units and extract key sentences based on the discourse structure. To detect the boundaries of the units, we make use of the pause information and linguistic information. We also incorporate another extraction method based on topic dependent keywords. We evaluate the proposed methods and their combination with 14 lecture transciptions. It is confirmed that the use of section boundary information and its combination with keyword-based method are effective.

keyword　lecture, discourse markers, topic words, automatic extraction, key sentence

1　緒論

近年の計算機性能の向上やメディア処理技術の進展に伴い，ディジタルアーカイブとして保存できる環境が整ってきている．しかし，音声のディジタルアーカイブはテキストとは異なり，そのままでは目的とする情報を迅速に検索し，短時間で全体の内容を把握することが困難である．したがって，内容を把握する上で，重要箇所や賛成・反対などの意見，話者情報などの 2 次情報をアーカイブに付与することが必要不可欠である．また自動処理の結果が不完全であっても，検索には十分である可能性は高く，人手による修正をあわせても効率的と期待される．

このような背景に基づいて，本稿では，講演を対象としてアーカイブ化に必要な重要文抽出を行う．その際，文章の再構成を行って要約するのではなく，重要と思われる文をそのまま抽出する．これには，抽出された文が日本語として自然であり，抽出された文を時間情報に基づいて音声と対応を取れるといった利点がある．

本研究では，講演の中でも学会講演を対象とする．一般に学会講演では，話題を問わずいくつかの論点が論理的に順序付けられて展開していくので，全体の話題構造のパターンが集約され，その境界が比較的明確である．そのため意味上の段落（以後これをセクションとよぶ）に分割し，セクションの冒頭と最後の文に着目する．ただしテキストにおけるセクション境界では改行や字下げなどの明示的な情報があるのに対して，話し言葉においてはそれらに相当するものはない．そこでポーズ情報および言語的情報を用いてセクション境界を検出することを考える．また，話題と関連のある重要な単語は当該講演において繰り返し出現すると仮定し，複合名詞を含めた名詞の統計情報も利用する．これらのセクション境界を用いた手法と話題語に基づく統計情報を用いる手法とを統合し，重要文を抽出する．これらの手法を「日本語話し言葉コーパス（CSJ）」[1] の学会講演を用いて評価した結果を報告する．

2　談話標識・話題語に基づく重要文抽出

学会講演は，大きく緒論，本論，結論の 3 つに分けられる．緒論については背景と目的に，本論については多くの場合，手法の説明と実験・評価に分けることができる．これは講演においては 1〜数枚程度のスライドに対応するまとまった話題の単位に該当し，これをセクションと定義する．そのセクションを基にした話題構造が存在し，その先頭もしくは末尾に重要文があると考えられ，これらの文を抽出することを試みる．

また各講演の話題に特有の語（話題語）を多く含む文が重要文であると仮定し，それらの文を抽出することを試みる．

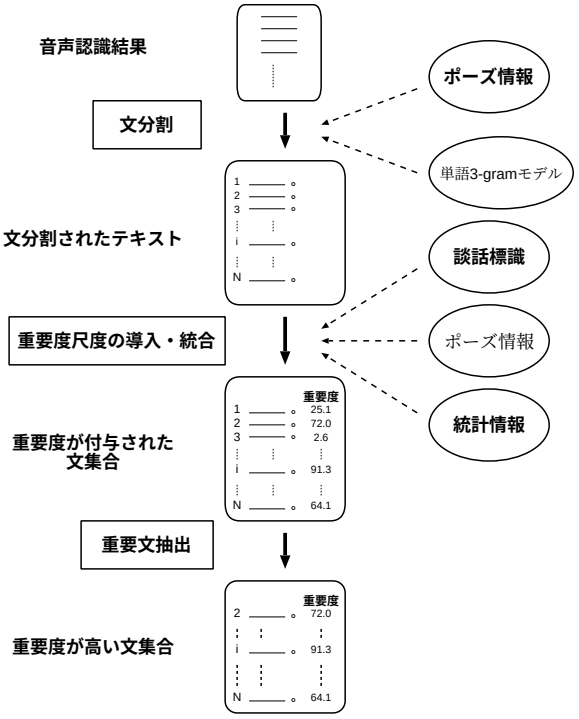

図 1: 重要文抽出の処理の概要

2.1　処理の概要

全体の処理の流れを図 1 に示す．まず，講演の書き起こしを文に分割する．次に分割された各文に対し，談話標識に基づいた手法，ポーズ情報を利用した手法，話題語に基づく統計情報を用いた手法の 3 つの評価尺度を用いて重要度を付与する．最後に各文の重要度をもとに，重要文を抽出する．これらの処理を以下に説明する．

2.1.1 文単位への分割

日本語の話し言葉においては，文の定義・境界が曖昧である．実際に CSJ の講演の書き起こしにはポーズ情報は付与されているが句点はない．それを用いて学習した言語モデルを用いて音声認識を行った結果にも句点は含まれない．そこで文を抽出するために，ここではポーズ情報及び単語 3-gram モデルを用いて文境界を検出し，句点を挿入する．

本稿では，文と文との間にはポーズが挿入されると仮定し，各講演毎のポーズの平均の長さを閾値として閾値以上のポーズが挿入された箇所について言語尤度を用いて句点を挿入かを判断する．

句点の有無による言語モデル尤度の差異に基づいて判定する手法は，ポーズが含まれる部分の前 2 単語"w_1w_2"と，後ろ 2 単語"w_3w_4"を取り出した上で，句点が入っていない状態の 4 単語をそのまま並べた単語列"$w_1w_2w_3w_4$"の尤度を単語数で正規化したパープレキシティ$-\frac{1}{4}\log P(w_1,w_2,w_3,w_4)$と，ポーズ部分に句点が挿入された単語列"$w_1w_2$句点 w_3w_4"のパープレキシティ$-\frac{1}{5}\log P(w_1,w_2,$句点$,w_3,w_4)$ を計算し，比較する．前者のパープレキシティが後者の 3 倍以内の値であれば，句点を挿入しない．

パープレキシティ計算用の言語モデルには，句点が付与されている Web 講演録（81 講演，1692802 形態素）で学習した単語 3-gram モデル（語彙サイズ 37462 語）を用いる．

2.1.2 重要度の計算と統合

各文 s_jに対し，話題の転換点を示す位置情報に基づく重要度と話題語の出現頻度に基づく重要度（$S_{KW(s_j)}$）を計算する．位置情報に基づく重要度尺度は談話標識から求める値（$S_{DM(s_j)}$）とポーズ長から求める値（$S_{pause(s_j)}$）を用いる．話題語に基づく重要度としては tf*idf 値を用いる．各重要度尺度については，次節以降で説明する．式 (1) に示す通り，線形重みづけにより統合し，これを文 s_jの重要度 S_{s_j}とする．

$$S_{s_j} = \sum \alpha_{choice} * S_{choice(s_j)} \quad (choice = pause, DM, KW) \tag{1}$$

ここで，重み$\alpha_{choice}(choice = pause, DM, KW)$ は 0 から 1 の間の値を取り，$\alpha_{pause} + \alpha_{DM} + \alpha_{KW} = 1$ の制約を満たすものとする．

2.1.3 重要文の抽出

前節の式 (1) により各文の重要度 S_{s_j}を求める．式 (2) で示される抽出率の範囲で重要度 S_{s_j}が高い順に抽出する．

$$\text{抽出率} = \frac{\text{1 講演からの抽出する文の数}}{\text{1 講演の総文数}} \tag{2}$$

2.2 談話標識の単語頻度および文頻度に基づく重要度尺度

セクション境界を検出する際には，長谷川らの手法 [2] を採用する．この手法は，各セクションの最初の一文に特徴的に現れる話題に独立な談話標識を抽出することにより，セクション境界を検出するもので，音声認識結果に対しても比較的頑健であると報告されている．具体的には次の統計量を用いる．

$$S_{DM(m_i)} = tf_{m_i} * \log\left(\frac{N_s}{sf_{m_i}}\right) \tag{3}$$

名詞 m_iの単語頻度 tf_{m_i}は，ポーズ長の平均値で定義する閾値以上のポーズの後の文集合において名詞 m_iが出現する回数である．文頻度 sf_{m_i}は，学習セットの全講演のすべての文で名詞 m_iの出現する文の数である．N_sは全講演における文の総数である．ある名詞 m_iについて，tf_{m_i}の値が大きいということはセクション境界の先頭部分に頻出する，つまり話題転換点で頻出している単語であることを表し，sf_{m_i}の値が小さいということは，多くの文にまんべんなく出現しないことを表す．

各文に出現するすべての談話標識に対するこの評価値 $S_{DM(m_i)}$の合計はセクション境界らしさを表す．$pause\_thres$ は閾値以上のポーズ長が存在するかを表し，0/1 の値をとる．

$$S_{DM(s_j)} = \sum_{m_i \in s_j} S_{DM(m_i)} * pause\_thres \tag{4}$$

2.3 ポーズ情報のみを用いた重要度尺度

言語的情報を用いない境界尤度も考える．セクション境界部分においては他の部分に比べて長いポーズが置かれると仮定し，各文の区切りのポーズ長のみから，セクション境界尤度を求める．具体的には，各文について，前後のポーズ長のうち値が大きい方をその文が持つポーズ長として定義し（式 (5)），平均と標準偏差で正規化した値を各文の境界らしさとし，

これを重要度尺度とした．ここで $SP(s_i)$ は文 s_iの直前のポーズ長を，n は各講演の文の数である．

$$pause(s_j) = \max(SP(s_j), SP(s_{j+1})) \quad (5)$$

$$S_{pause(s_j)} = \frac{pause(s_j) - \mu}{\sigma} \quad (6)$$

$$\mu = \frac{\sum_j SP(s_j)}{n}, \quad \sigma = \sqrt{\frac{\sum_j SP(s_j) - \mu}{n}}$$

2.4 話題語を考慮した重要度尺度

話題語の重要度尺度として名詞を対象にした tf*idf 値による統計的尺度を用いる [3, 4]．ただし，数詞についてはその対象外とした．この際に，単純に基本的な名詞を選ぶのではなく，連続して出現する名詞列を複合語として扱う．例えば，ある講演で「音声」「認識」という 2 単語が連続して出現したとき，これらを複合名詞「音声認識」の 1 単語とみなす．

単語 w_iの tf*idf 値は式 (7) により定義される．

$$KW_{w_i} = tf^a_{w_i} * \log\left(\frac{N_d}{df^b_{w_i}}\right) \quad (7)$$

tf_{w_i}は話題語の名詞 w_iの 1 講演内での出現回数を表わす．df_{w_i}は名詞 w_iが出現する講演数を表し，全講演数 N_dをこれで除したものが idf 値である．a，b はそれぞれ，tf 値，df 値の重みであるが，本稿ではともに 1 とした．ある名詞 w_iについて，tf 値が大きいと，その講演で頻出している単語であることを表し，df 値が大きいと話題に関係なく，多くの講演にまんべんなく出現していることを表す．

各文について含まれる名詞の tf*idf 値から重要度を求める．総和を各文の重要度とする方法 (式 (8)) と，一名詞あたりの平均を各文の重要度とする方法 (式 (9)) の 2 通りの方法で実験を行う．ただし，$n(s_j)$ は文 s_jに含まれる名詞数を表す．

$$S_{KW(s_j)_{total}} = \sum_{w_i \in s_j} KW_{w_i} \quad (8)$$

$$S_{KW(s_j)_{average}} = \frac{\sum_{w_i \in s_j} KW_{w_i}}{n(s_j)} \quad (9)$$

3 CSJ の講演を用いた評価実験

3.1 学習・評価データ

学習データには，CSJ の学会講演（688 講演）の書き起こしを用いる．形態素解析システム ChaSen ver2.02[5] を用いて品詞情報が付与された形態素から名詞を選んだ．評価データには 14 件の学会講演 [1] の書き起こしを用いる．正解となる重要文は人手で選んだ．全体の文の数に対する重要文の数の割合は 21.6%であった．評価の尺度には，再現率 (Recall)・適合率（Precision）・F 値（F-measure）を用いる．F 値は以下の式で表される．

$$F\text{-}measure = \frac{2 * Recall * Precision}{Recall + Precision} \quad (10)$$

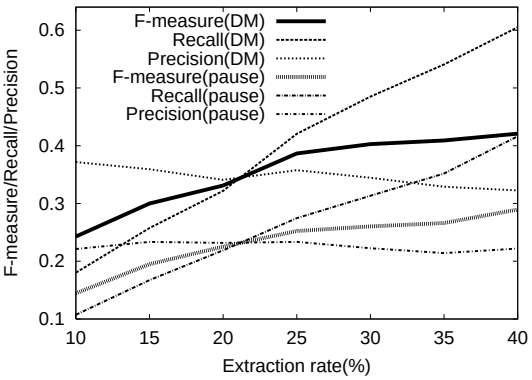

図 2: 談話標識に基づく重要度尺度 (DM) およびポーズ長を用いた重要度尺度 (pause) による重要文抽出精度

3.2 セクション情報を用いた重要文抽出

2.1.1 節で述べた手法により文境界が与えられたテキストに対して，提案するセクション情報を用いて抽出率 30%で重要文抽出の実験を行った．その場合の再現率を表 1 に示す．比較のため，人手によりセクションに分割し境界前後の文を一定数抽出する方法，セクション情報を考慮せず 1 講演全体における冒頭および末尾から一定の文章を抽出する方法によって得られた抽出率 30%の結果も示している．談話標識（式 (4)）およびポーズ長のみ（式 (6)）を用いた重要文抽出の結果を図 2 に比較する．セクション情報を用いた場合の方が，用いない場合よりよい結果を示した．つまり重要文を抽出する上で，セクション情報を用いることが有効であることが示された．また，ポーズ情報のみで抽出で行うよりも談話標識を用いて抽出を行ったほうが抽出精度が高い値を示した．これは，話者によりポーズの長さ，入れ方が異なるため精度が低くなったと考えられ，ポーズ情報のみでの抽出には限界があると考えられる．

[1] 文献 [2] と基本的に同じ評価セット

表 1: 位置情報を用いた重要文抽出（抽出率 30%）

セクション情報	用いる			用いない
	人手	談話標識	ポーズ長	
再現率 (%)	54.2	48.5	31.3	27.5

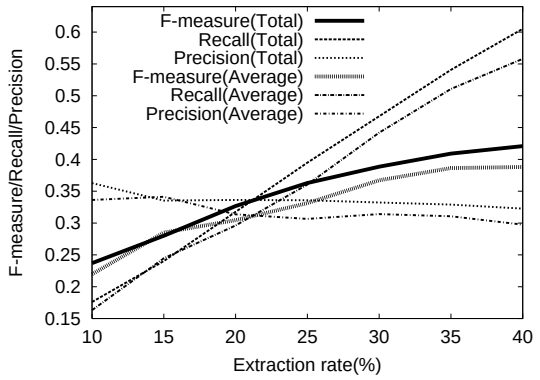

図 3: 各話題語の tf*idf 値の総和（Total）及び平均（Average）を文重要度としたときの重要文抽出精度

3.3 話題語の統計情報に基づく重要文抽出

各文に含まれる話題語の統計情報 tf*idf 値の総和（式 (8)）と平均（式 (9)）をその文の重要度とする方法の比較を行う．F 値，再現率，適合率を図 3 に示す．総和を重要度尺度とする方法が，平均を重要度尺度とする方法に比べて若干精度が高い．これは比較的長めの文が重要文となっていることを示す．本研究では文の数に対して抽出率を設定しているため，同じ量の文を抽出した際に，名詞の数が多く情報量の多い長めの文が抽出されやすいためであると考えられる．しかし，文の数ではなく文字数に応じて抽出率を設定した際には，tf*idf 値の平均を重要度尺度とする方が精度が高くなる可能性がある．

3.4 重要度の統合

これまでに挙げた談話標識に基づく重要度尺度とポーズ長に基づく重要度尺度と話題語に基づく重要度尺度の 3 つの重要度を統合して新たな尺度として実験を行った．結果を図 4 に示す．

単独で求めた重要度尺度では，談話標識に基づく重要度尺度が最も高い精度を得た．話題語に基づく重要度により抽出した結果は，談話標識に基づく重要度により抽出した結果に比べてやや精度が低いものの，ほぼ同程度の結果が得られた．ポーズ長を用いた重要度により抽出した結果は，最も低い精度であった．

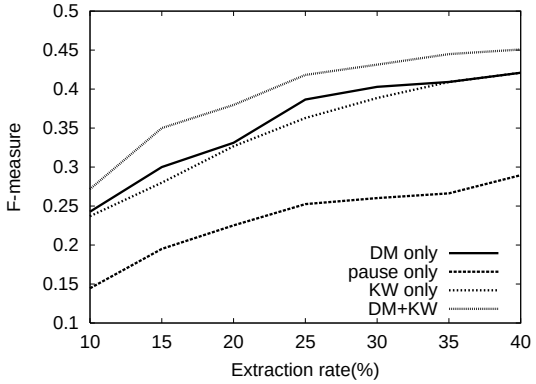

図 4: 談話標識に基づく重要度尺度（DM），ポーズ長を用いた重要度尺度（pause），話題語に基づく重要度尺度（KW）の統合による抽出精度

表 2: 3 重要度尺度の統合による抽出精度（抽出率 30%）

重要度尺度	再現率	適合率	F 値
DM	48.5%	34.5%	0.403
pause	31.3%	22.3%	0.260
KW	46.8%	33.2%	0.389
DM+pause	45.5%	32.3%	0.378
DM+KW	**51.9%**	**36.9%**	**0.431**
pause+KW	45.5%	32.1%	0.378
DM+pause+KW	51.5%	36.6%	0.428

各々単独の場合，ポーズ長に基づく重要度尺度以外の 2 つの重要度尺度を組み合わせた場合，3 つの重要度尺度を組み合わせた場合の結果（F 値）を図 4 に示す．式 (1) にしたがって線形の重みづけ和により行う．3 つの重要度の混合重みは F 値が最大となるように最適化している．また，抽出率 30%における各手法単独の精度と，統合した場合の精度を表 2 に示す．精度の低いポーズ長に基づく重要度尺度を統合した場合は，基本的に統合による効果は見られない．これに対して，談話標識に基づく重要度尺度と話題語に基づく重要度尺度とを統合した場合は各々単独の場合より抽出精度が改善し，抽出率 30%において，再現率 51.9%，適合率 36.9%，F 値 0.431 の結果が得られ，最高の精度となった．このときの混合重みαの値（混合比）は，$\alpha_{DM} : \alpha_{KW} = 0.3 : 0.7$ であった．

3.5 音声認識結果への適用

音声認識は，大語彙連続音声認識エンジン Julius rev3.2[6]（逐次デコーディング）を用いて行う．その際，音響モデルには CSJ の学会講演男性話者 60

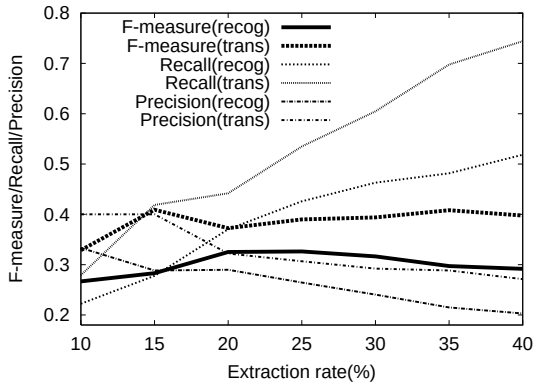

図 5: 音声認識結果 (recog) と書き起こし (trans) からの重要文抽出の精度

表 3: 音声認識結果からの重要文抽出（抽出率 30%）

	再現率	適合率	F 値
音声認識結果	25/54 (46.3%)	25/104 (24.0%)	0.316
書き起こし	27/43 (62.8%)	27/89 (30.3%)	0.448

時間で学習した PTM triphone モデル [7] を，言語モデルには CSJ の学会講演と模擬講演（2.7M 単語）から学習した単語 3-gram モデルを用いる．

音声認識結果を用いた評価は，書き起こしを用いた場合の実験でも使用した A01M0007，A01M0074，A03M0100 の 3 件の学会講演を用いて行った．このときの抽出結果を図 5 に示す．図 5 には書き起こしを用いた場合との比較も示している．また抽出率を 30%にした場合の結果を表 3 に示す．音声認識結果を対象にした場合は書き起こしを対象にした場合に比べて，大きく精度が低下した．

4 結論

実際の学会講演の書き起こし及び音声認識結果から重要文を抽出する手法を提案した．本研究ではセクションという単位を設定し，その境界前後に重要文が集中して存在しているという仮定に基づいて，それらを抽出する手法を提案した．その際，話し言葉においてはセクション境界が明示的でないため，セクション境界検出には談話標識を利用する方法とポーズ情報のみを用いる方法を検討，比較した．また話題語を多く含む文が重要文であるという仮定に基づく tf*idf 法も導入した．

セクション境界情報を用いることで，講演全体の位置情報のみに基づく単純な手法よりも高精度に重要文抽出を行えることを示した．セクション境界の抽出には談話標識に基づく手法がポーズ長のみによる手法より優れていることも確認した．談話標識に基づく重要度と話題語に基づく重要度を統合することにより相乗効果が得られ，再現率 51.9%，適合率 36.9%，F 値 0.431 となった．音声認識結果へ適用したところ，再現率 46.3%，適合率 24.0%，F 値 0.316 であった．今後は，セクション分割精度の改善を図るとともに，認識誤りに対処できる枠組みを検討していく予定である．

参考文献

[1] 古井貞熙, 前川喜久雄, 井佐原均. 科学技術振興調整費開放的融合研究推進制度－大規模コーパスに基づく「話し言葉工学」の構築－. 日本音響学会誌, Vol.56,No.11,pp.752-755, 2000.

[2] 長谷川将宏, 秋田祐哉, 河原達也. 談話標識の抽出に基づいた講演音声の自動インデキシング. 情報処理学会論文誌, Vol. 43, No. 7, pp. 2222–2229, 2002.

[3] 伊藤山彦, 松本賢司, 谷田泰郎, 柏岡秀紀, 田中英輝. 講演文を対象にした重要文抽出実験.「話し言葉の科学と工学」ワークショップ予稿集, pp157-164, 2001.

[4] 野畑周, 関根聡, 内元清貴, 井佐原均. 話し言葉コーパスにおける文の切り分けと重要文抽出.「話し言葉の科学と工学」ワークショップ予稿集, pp93-100, 2002.

[5] 松本裕治, 北内啓, 山下達雄, 平野善隆, 松田寛, 浅原正幸. 日本語形態素解析システム 茶筅 version 2.0, 12 1999.

[6] 河原達也, 加藤一臣, 南條浩輝, 李晃伸. 話し言葉音声認識のための言語モデルとデコーダの改善. 2001.

[7] 南條浩輝, 加藤一臣, 李晃伸, 河原達也. 大規模な日本語話し言葉データベースを用いた講演音声認識. 電子情報通信学会論文誌, Vol. J86-D-II, No. 4, pp. pp.450–459, 2003.