音声言語情報処理 35－4
（2001．2．2）

ニュース音声の話題決定のためのワードスポッティングの評価

加藤　大知　　山下　洋一
立命館大学理工学部情報学科
〒525－8577　滋賀県草津市野路東1－1－1
E-mail：taichi@slp.cs.ritsumei.ac.jp, yama@cs.ritsumei.ac.jp

概要：ニュース音声などの連続音声の話題を自動決定するために、ワードスポッティングを用いてキーワードを抽出し、抽出されたキーワード列に基づいて話題を決定する手法を検討する。本報告では、話題決定に用いるキーワードセットの決定法を、ワードスポッティングにおける検出率/湧き出し誤り率および話題決定率の観点から評価する。キーワードの選出尺度として、話題との関連性を示す尺度として相互情報量、負の値を持つ X^2値、キーワードの検出のしやすさの尺度として湧き出し誤りの起こり難さ（DF 値）を用い、これらを組み合わせることによりキーワードセットを選出した。
キーワード：ワードスポッティング、湧き出し誤り、相互情報量、負の値を持つ X^2値

Evaluation of Word Spotting for Topic Identification of News Speech.

Taichi Kato,　Yoichi Yamashita.
Department of Computer Science, Ritsumeikan University
1-1-1 Noji-Higashi, Kusatsu-shi, Shiga, 525-8557 Japan

Abstract : This paper describes a method for automatically identifying topic for speech materials such as news speech. Topic identification is based on a sequence of keywords extracted by a word spotting technique. We used three types of measures of keyword selection, mutual information, modified X^2 score, and detection facility. The first two measures indicate degree of relevance to topic, and the last one estimates the number of false alarms of a word. Combination of these mesures selected several keyword sets. The keyword sets were evaluated from viewpoints of performance of the keyword detection and the topic identification.
Keyword : word spotting, false alarm, mutual information, X^2

1　はじめに

現在、社会の情報化が進み、様々な情報が大量に蓄積されるようになってきた。蓄積される情報が増えれば増えるほど、必要な情報だけを効率よく取り出すことが難しくなり、効率の良い検索手法や、蓄積されたデータのインデキシングが必要となってくる。例えば、テレビなどのニュース音声のデータベースから必要な記事を検索するときには、記事の話題分類が有効なインデックスとなる。音声メディアに対する検索を容易にするために、ニュースなどの音声データの話題（ジャンル）を自動的に決定する方法が、近年広く研究されている［1］［2］［3］。

人が話の内容（話題）を決定するときに話に出てくる単語が重要な情報となっている。連続音声の話題を決定する時にも、音声中より話題との関連性が高い語（以下、キーワード）を抽出し、その情報を用い話題決定を行うことは、有効な手段であると考えられる。

キーワードの抽出方法は、音声中からワードスポッティングによって直接抽出する方法と連続音声認識を行い、認識結果から必要なキーワードを抽出する方法などがこれまでに報告されている。ワードスポッティングによるキーワード抽出は、連続音声認識に比べ、言語モデルと対象データのずれ、話題決定に有効な新語の扱い（未知語）、いいよどみへの対応の問題がないことから、本研究ではワードスポッティングによるキーワード抽出を行っている。

ワードスポッティングに基づいて話題決定を行う場合には、話題への関連性だけでなく検出のしやすさも考慮して検出すべきキーワードセットを決定しなければならない。著者らは、これまでに音声認識シミュレーションによってワードスポッティングによる単語の検出容易さを見積もる手法を検討してきた［4］［5］。本研究では、検出のしやすさも考慮したいくつかのキーワード選択手法を試み、ワードスポッティング性能話題決定率の観点から評価する。

2　ワードスポッティングに基づいたニュース音声の話題決定

図１に示すように、1)予めキーワードセットを選択しておき、2)ニュース音声からキーワードの検出を行い、3)検出されたキーワード列から各話題の話題らしさを算出して、最も値の高い話題を取り出し、話題決定を行う。検出されたキーワード列 W_i に対して式(1)を用いて、話題 T_j の話題らしさ $S(T_j)$を算出する。

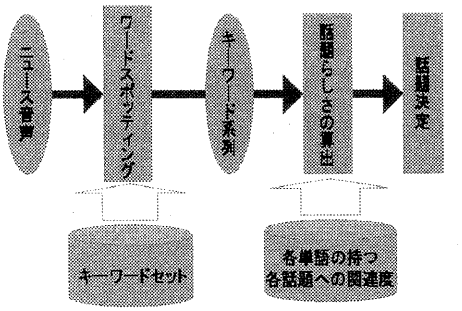

図１　話題決定の処理手順

$$S(T_j)=\sum_i(\log P(w_i|T_j)+c*S(w_i)) \qquad (1)$$

c：$S(w_i)$の重み

$P(w_i|T_j)$：話題T_jにおける単語w_iの出現確率

$S(w_i)$：音声中で単語w_iが検出されたときの尤度

3　キーワード選出の尺度

キーワードの良さとして

1）　相互情報量
2）　負の値を持つ x^2 値
3）　湧き出し誤りの起こり難さ（DF 値）

の３つを考え、これらを組み合わせて利用することによりキーワードを選択する。キーワードの良さを示す尺度として、下記の１～６の手法について評価を行う。

手法１－１	相互情報量
手法１－１	相互情報量×DF 値
手法２－１	x^2値
手法２－２	x^2値×DF 値
手法２－３	x^2値＋ω×DF 値(ω=10)
手法２－４	x^2値＋ω×DF 値(ω=50)

まず、毎日新聞 CD－ROM４５ヶ月分（１９９１年～１９９４年９月）に出現する名詞で、出現数上位２０，０００単語を選出する。選出された２０，０００単語に対し上記の手法１から６の方法でスコアを算出し、スコアの

高い、４音素以上の単語を各３,０００単語ずつ選択し使用する。

３．１　相互情報量

単語wと話題Tの相互情報量を考える。Tを記事の話題を表す確率変数、Wを記事中のある単語を表す確率変数とすると、相互情報量I(T;W)は、

$$\begin{aligned} I(T;W) &= I(W;T) \\ &= H(W) - H(W|T) \\ &= \sum_{w}[-p(w)\log_2 P(w) \\ &\quad + \sum_{T} p(T)p(w|T)\log_2 p(w|T)] \end{aligned} \quad (2)$$

となり、様々な単語から寄与を合計したものとなる。従って、式(2)の［ ］内の値

$$G(w) = -p(w)\log_2 p(w) + \sum_{T} p(T)p(w|T)\log_2 p(w|T)$$

が大きい単語wほど特定の話題に多くの情報を持っていると考える［１］。

３．２　負の値を持つX^2値

式（３）に示す負の値を持つX^2値は、分野における単語の偏りを示す指標として用いることができる。単語wの出現確率は全分野に対して等しいという仮説を設定し、この仮定に基づいて単語 w の分野における予測頻度 m を計算する。また各単語wについて分野tにおける頻度xを求める。もし負の値を持つX^2値が十分に大きな値になれば特定の分野に偏って表れる単語ということになり、分野識別に有効な単語とみなせる［２］［３］。

$$m_{ij} = \frac{\sum_{j=1}^{n} x_{ij}}{\sum_{i=1}^{m}\sum_{j=1}^{n} x_{ij}} \quad (3)$$

$$X_{ij} = \frac{(x_{ij} - m_{ij})\,|\,x_{ij} - m_{ij}\,|}{m_{ij}} \quad (4)$$

m：異なり単語数　　n：分野数

x_{ij}：単語 w_iの分野 t_jにおける頻度

m_{ij}：単語 w_iの分野 t_jにおける予測頻度

３．３　湧き出し誤りの起こり難さ（DF値）

ワードスポッティングでは検出すべき単語が長いほど検出が容易で誤検出も少なくなる。しかし、検出の容易さは単語の音素数（モーラ数）だけでなく単語を構成する音の並びにも依存すると考えられる。そこで、単語認識シミュレーションを行い、湧き出し易さを見積もることを考える[４][５]。

ワードスポッティングで検出すべき単語wの音素長を L、wを構成する音素の並びを$pw_1, pw_2, \cdots, pw_L$とする。文における音素列の生成をｎ重マルコフ過程によってモデル化し、シミュレーションによって長さMのK個の音素列s_k（k＝1,・・・,K）を生成する。s_kを構成する音素の並びを$ps_{k1}, ps_{k2}, \cdots, ps_{kM}$とする。音素$p_i$が音素$p_j$に認識される確率、音素$p_i$が脱落する確率、音素の挿入があるときに挿入される音素がp_jである確率をそれぞれ、$S(p_j \mid p_i)$、$D(p_i)$、$I(p_j)$で与えられたとする。ただしNを音素の種類としたとき

$$\sum_{j=1}^{N} S(p_j \mid p_i) + D(p_i) = 1, \sum_{j=1}^{N} I(p_j) = 1 \quad (5)$$

である。ここでs_kの部分音素列と単語wとの様々な対応関係を考える。その中で、s_kの部分音素列が単語wと認識される最大の確率 MP_kを終点フリーの動的計画法を用いて求める。s_kの部分音素列としては、長さがL−DかL+Dの範囲を考える。したがって、M.L+Dでなければならない。MP_kは、

1) $g(1,1) = S(pw_1 \mid ps_{k1})$,
$g(1, j) = 0 \quad (j = 2, \cdots, L)$
2) $i = 2, \cdots, L + D$に対して
3),4),5)を繰り返す。
3) $j1 = \max(1, i - D),\ j2 = \min(L, i + D)$
4) $j = j1, \cdots, j2$に対して5)を繰り返す。
5) $$g(i,j) = \max\begin{cases} g(i-1, j-1) \times S(pw_j \mid ps_i) \\ g(i-1, j) \times D(ps_i) \\ g(i, j-1) \times I(pw_j) \end{cases} \quad (6)$$
6) $MP_k = \max(g(L - D, L), \cdots, g(L + D, L))$

で与えられる。ここでDは非線形伸縮の程度を制御するパラメータである。さらに湧き出し誤りの起こりやすさFPは、K個の音素列についてMP$_k$の和をとったものとして、

$$FP = \sum_{k=1}^{K} MP_k \qquad (7)$$

となる。さらに、湧き出しの起こり難さを、

$$DF = -\log FP \qquad (8)$$

とする。

4　評価実験

ワースポッティングによって検出されたキーワードの持つ各話題への寄与度を用いて各話題の話題らしさを式（1）を用いて算出する。話題決定に用いる話題は、“国際”、“経済”、“家庭”、“文化”、“科学”、“芸能”、“スポーツ”、“社会”の８話題である。

4．1　学習、評価データ

各キーワードの各話題に対する寄与度の学習データとして、「毎日新聞 CD−ROM」４５ヶ月分('９１年１月〜'９４年９月)を用い、評価データには NHK１時のニュース２２日分７８記事（'９８年８月収録、総計１．８３時間、１記事平均約１００秒、名詞単語の種類は２４８６単語）を使用した。７８記事の内訳は、社会５３記事、国際１６記事、経済７記事、スポーツ２記事であった。

4．2　ワードスポッティング

まず、キーワードセットを比較する。表１に各キーワードセットの平均音素長、出現キーワードの平均音素長、出現のべ回数平均音素長、出現キーワードの種類、１話題に出現する平均キーワード数、評価データ中の単語におけるキーワードの割合(カバー率)を示す。

次に文献[４]に述べられている手法でワードスポッティングを行う。図２にしきい値を変えた時の各手法でのキーワード検出率を示す。ワードスポッティングのしきい値は、−１〜−２０で実験を行った。図２よりDF値を考慮した手法２－２のキーワードセットが湧き出し誤りの減少、より高い検出率を得ていることがわかる。手法１の中では、手法１－１よりも手法１－２によるキーワードセットの方が検出率が低かった。手法１－２はDF値を考慮したキーワード選択法であり、比較的長い音素長のキーワードが多く選択されていることが期待される。しかし、３０００キーワードの平均音素長では手法１－１より手法１－２の方が長いものの、実際に音声データに出現したキーワードの平均音素長では、表１に示すように手法１－２の方が短くなっている。このため手法１－２によるキーワードセットのスポッティング性能が低くなったと思われる。

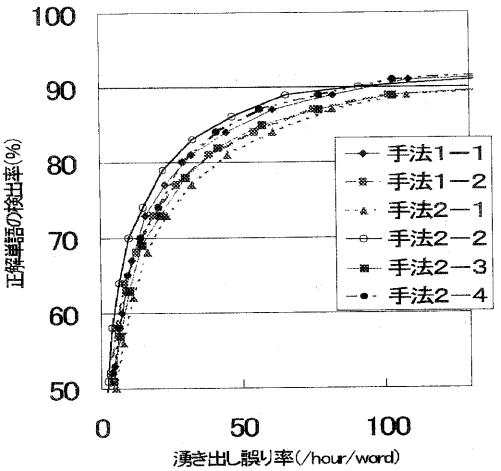

図２　正解単語の検出率

表１　評価テキストデータにおけるキーワードセットの比較

	手法 1-1	手法 1-2	手法 2-1	手法 2-2	手法 2-3	手法 2-4
キーワードセットの平均音素長	6.48	6.84	6.33	7.57	6.61	7.08
出現キーワードの平均音素長	6.04	6.02	5.79	6.36	6.00	6.08
キーワードの出現のべ回数	3326 回	989 回	3645 回	1811 回	3503 回	3435 回
出現キーワードの種類	710 単語	299 単語	756 単語	404 単語	711 単語	687 単語
1 話題に出現する平均キーワード数	8.5 単語	3.6 単語	9 単語	4.8 単語	8.5 単語	8.2 単語
カバー率	29%	12%	30%	16%	29%	28%

4．3　話題決定

まず、テキストデータを対象としてキーワード列に基づく話題決定を行った。これは、ワードスポッティングを行ったときに検出率が１００％で湧き出し誤りが１つも出なかった状態と同じである。その結果を図３に示す。DF 値を考慮した手法１－２、２－２、３、４が DF 値を考慮しない手法１－１、２－１より話題決定率がやや低い結果となった。この理由として、DF 値を考慮することによりキーワードと話題との関連性が下がってしまったことが考えられる。表１でのカバー率からわかるように、手法１－２ が極端に話題決定率が低いのは、評価データに出現するキーワードの割合が他に比べ少な過ぎたことが原因だと思われる。

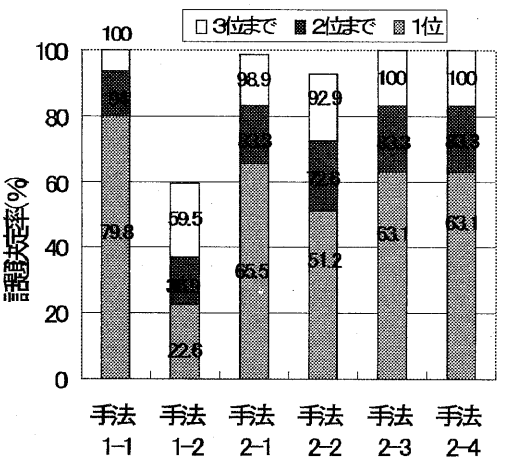

図３　テキストデータでの話題決定

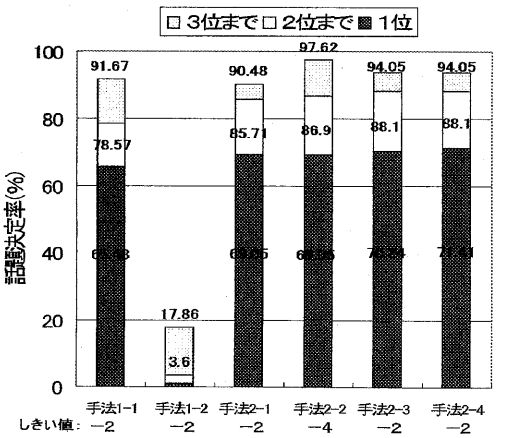

図４　各手法での話題決定率

次に図１で示した手順で実際に話題決定実験を行った。各手法で最も決定率の高かったしきい値での結果を図４に示す。手法１－２を除いて、どの手法も同じような話題決定率となった。表２に手法１－１による話題決定の結果を示す。他の手法でもほぼ同様の傾向を示した。これからわかるように、話題認識結果がほとんど社会に偏ってしまい、もともと評価データの社会記事が占める割合が約６９％であったため、どの手法も同じように７０％弱の話題認識率を得たと考えられる。DF 値を考慮することにより話題決定率の向上を期待していたが、図４のような結果となり、キーワード選択法の優劣をつけることができない結果となった。

次に表１に示すように、どの手法も出現キーワードの検出数が５５０から６５０単語（1 話題あたり約６～７単語）までの間で、最も高い話題決定率を得ている。このことからキーワード数（今回は３０００単語）が多すぎることが考えられる。キーワード数を減らせば、必然的に湧き出し誤り数も減少するので、より適正なキーワード数の考慮が必要である。参考までに各しきい値での出現単語の検出数を図５に示す。手法２－３と手法２－４は、ほぼ同じ結果となったので重なって表示されている。

表２　話題決定の結果(例：手法１－１)

	社	経	国	ス	科	芸	家	文
社	49	4	-	-	-	-	1	1
経	3	3	2	-	-	-	-	-
国	13	-	6	-	-	-	-	-
ス	2	-	-	-	-	-	-	-

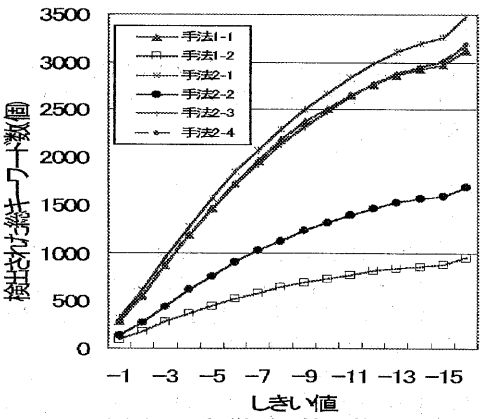

図５　正解単語の検出数

次に手法１－１ではテキストデータで約８０％の話題決定率が実際にスポッティングを行った実験で約６５％と下がってしまい、逆に手法２－２では、約５１％から６９％へと上がってしまっている。テキストデータでの実験と音声データでの実験の結果が同じ程度の話題決定率であれば、湧き出し誤りの影響はないと考えられるが、これらのことより湧き出し誤りの影響が大きいことがわかる。

5　終わりに

本研究では、ワードスポッティングに基づいた話題決定を行うときに用いるキーワードセットの決定法を、ワードスポッティングにおける検出率/湧き出し誤り率および話題決定率の観点から評価を行った。結果として、

- DF 値を考慮したキーワードセットを用いることによりスポッティング性能は向上した。
- 話題決定では DF 値を考慮したキーワードセットの有効性は見られなかった。
- 評価データに出現する単語のカバー率が低いと話題決定率も低くなった。

ことがわかった。このことより湧き出し誤りをいかに抑制するかということとワードスポッティング自体の性能向上が必要であると考えられる。

今後は、より良いキーワードセットの選出と、適正なキーワード数、湧き出し誤りの抑制の方法などを検討し、再実験を行う予定である。

謝辞
本研究では、キーワードセットの学習に毎日新聞 CD-ROM（'９１年１月～'９４年９月）を使用した。

参考文献

[1] 恒川 俊克、山下 洋一、溝口 理一郎、“キーワードスポッティングに基づく話題決定”、情処研報、SLP２０－１１、ｐｐ６１－６８（１９９８－２）.

[2] 鷹尾 誠一、緒方 純、有木 康雄、“ニュース音声の記事分類におけるキーワード選択法の比較”、情処研報、SLP２２－１５、ｐｐ７５－８２（１９９８－７）.

[3] K. Ohtsuka、T. Matsuoka 他、“TOPIC EXTRACTION MULTIPLE TOPIC-EORDS IN BROADCAST-NEWS SPEECH”、ICASSP98、ｐｐ３２９－３３２（１９９８）.

[4] 山下 洋一、“ワードスポッティングにおける湧き出し誤りの予測”音講論集、１－１－１１、ｐｐ２１－２２（１９９８－９）.

[5] 山下 洋一、“音声認識シミュレーションに基づくワードスポッティング精度の予測”、信学技法、SP９８－９５、ｐｐ６５－７０（１９９８－１１）.