ニュース検索タスクにおけるシステム要求と雑談の判別

佐古　淳[†]　田中　克幸[††]　滝口　哲也[††]　有木　康雄[††]

† 神戸大学大学院自然科学研究科　〒 657–8501 兵庫県神戸市灘区六甲台町 1–1
†† 神戸大学大学院工学研究科　〒 657–8501 兵庫県神戸市灘区六甲台町 1–1
E-mail: †{sakoats,yamagata}@me.cs.scitec.kobe-u.ac.jp, ††{takigu,ariki}@kobe-u.ac.jp

あらまし　情報網やデジタル化の発展に伴い，情報整理が困難な音声・映像メディアが増大している．ユーザーが欲しい情報を簡単に得るための仕組みが必要である．誰もが簡単に利用可能な仕組みとして，音声インターフェイスによる情報検索システムが有望であると考えられる．音声をインターフェイスとして用いる際，システムに対してなされた発話か，周りの人間に対してのものかを判別する必要がある．この問題に対し，柔軟な発話を受理可能なものとして，音声認識結果をブースティングによってシステム要求か雑談かを判別する手法の提案を行ってきた．本稿では，検索キーワードを含むためにコマンドが一定でない情報検索タスクにおいて，ブースティングによるシステム要求検出の有効性を検証する．検索キーワードをクラス化しブースティングを行うことで適合率 0.81，再現率 0.85，F 値 0.83 を実現した．
キーワード　システム要求判別, 情報検索, 動画, ブースティング, 音声認識

System Request Discrimination for Information Retrieval

Atsushi SAKO[†], Katsuyuki TANAKA[††], Tetsuya TAKIGUCHI[††], and Yasuo ARIKI[††]

† Guraduate School of Science and Technology, Kobe University
Rokkodaicho 1–1, Nada-ku, Kobe, Hyogo, 657–8501 Japan
†† Guraduate School of Engineering, Kobe University
Rokkodaicho 1–1, Nada-ku, Kobe, Hyogo, 657–8501 Japan
E-mail: †{sakoats,yamagata}@me.cs.scitec.kobe-u.ac.jp, ††{takigu,ariki}@kobe-u.ac.jp

Abstract　The advancement of information technology, which includes such developments as Web2.0, on digital TV and Broadband, enables anyone and everyone to access and participate to access any form of media, like documents, movies, images etc via the internet very easily. However, due to information growth and the decentralization of platforms, users are faced with increasing difficulty in finding the information that they really are interested in. Speech user interfaces promise useful system of information retrieval for every user. It is necessary to discriminate system requests from human-human conversation speeches for speech user interfaces. We had proposed the boosting method that discriminates system requests from chats based on 1-best result of speech recognition system. This method can retrieve various expressions due to boosting algorithm. In this paper, we ascertain whether the boosting method is efficient for information retrieval task whose commands are nonconstant due to including search keywords. As a result of grouping search keywords into a class, the experimental results showed that 0.81 of precision, 0.85 of recall and 0.83 of F-measure.
Key words　System Request Detection, Information Retrieval, video, Boosting, Speech Recognition

1. はじめに

近年，ブロードバンド化やユビキタス化などインターネットの利用環境や情報技術の発展に伴い，簡単に情報収集・提供が可能となっている．インターネットの速度向上に従って流れる情報はより多くなる．近年では，テキストだけではなく画像・音声・映像など，あらゆるメディアの情報が普及し，WWW は大規模なマルチメディ

ア情報源となった．そのため，このような大量の情報からユーザーの欲しい情報のみを検索する仕組みが必要とされている．音声インターフェイスを用いた情報検索システムは，誰もが簡単に利用できるシステムとして有望である．従来，我々は音声インターフェイスを用いた動画検索システム NetNews [1] について提案を行ってきた．しかし，従来の NetNews ではシステムが理解可能なコマンド以外が入力されると，コマンドの湧き出しが発生し，誤動作を引き起こすという問題があった．そのため，複数人で視聴する際の雑談が問題となっていた．

また一方で，我々は，ロボットやカーナビを対象としたタスクにおいて，システムへの要求と雑談とを判別する手法について提案を行ってきた [2]．提案手法では，音声認識結果に対してブースティングを用いることで，システム要求発話の柔軟性・多様性を保持したまま自動的に区別を行うことが可能であった．

本稿では，インターネット上の動画検索システム NetNews に対するブースティングを用いたシステム要求検出の適用について述べる．本研究で取り扱う動画検索タスクは，従来のロボット・カーナビタスクと比べ，検索のためのキーワードがコマンドに含まれるため，コマンドが一定ではない点が異なっている．

以下，次章では本研究で用いた動画検索システムについて述べ，3. 章で提案手法について述べる．4. 章で評価実験と考察について述べ，5. 章でまとめる．

2. 動画検索システム NetNews

本章では，本研究で用いた動画検索システム NetNews の概要について述べる．図 1 に NetNews の構成を示す．ネット上にあるニュース動画を自動的に入手するとともに，時事のトピックをキーワード音声検索により観覧することが可能である．また，表 1 に NetNews の音声ユーザーインターフェイスの仕様を示す．

時事のトピックを音声により検索する際の問題点として，単語辞書の問題があげられる．辞書がコンパクトな場合，一般的に認識性能の向上が期待できる．しかしその反面，未知語の増加を招き，特に新しいトピックのキーワードが辞書から抜け落ちてしまう危険性が高まる．一方，大きな辞書を構築すると，未知語は減らすことができるが，認識性能の低下を招く．また，認識は出来たものの，認識結果で検索しても動画は見つからないといった状況も考えられる．すなわち，検索対象の動画に対して適切な内容・サイズの辞書を構築する必要がある．NetNews では，日々，クローリングを行い WWW 上の動画のインデックスを作成している．このとき，動画のインデックスを音声認識辞書としてダイナミックに登録することにより，適切なサイズの辞書を維持することが可能となっている．

NetNews では，以下のような手順により，動画のインデックスを作成している．

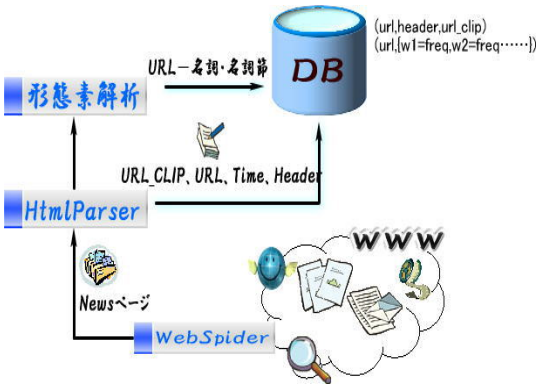

図 1　動画検索システム NetNews の概要
Fig. 1　Overview of the video search system: NetNews

表 1　NetNews における音声インターフェイスの仕様
Table 1　Abilities of the video search system: NetNews

機能	ニュース映像の検索 映像の早送り・巻き戻し トップニュースの表示
コマンド例	○○のニュース ○○って何 次・前の映像 最新情報リスト

- WWW を巡回し，ニュースサイトからヘッダーリンクを探して（url,header）のペアリストを作る．
- そのリンクを辿って詳細記事ページを集める．集めた各詳細記事ページに html パーザをかけ，記事部分と動画部分を切り出す．
- html タグ内から動画ＵＲＬを探し出して抜き出す．
- リンクで集めたヘッダーを探してそれを記事の始まりとし，時間の表示などのタイムスタンプをもとに記事の終わりと判断して記事を切り抜く．url と詳細記事を用いて動画（url_clip）のインデックス化を行う．
- 抜き出してきた詳細記事とヘッダーを文単位に分解して，chasen [3] により形態素解析を行う．
- 解析結果から，名詞，未知語を取り出してインデックスキーとする．また，名詞の連続とカタカナの連続は，１つのフレーズとしてキーとみなす．

この結果，(url,header,url_clip)，(url,w1=freq1,w2=freq2……) のインデックステーブルが作成され，キーワードによる検索が可能となる．

NetNews を複数人で使用しながら収録を行った．収録用のマイクはユーザーの胸元に設置した．総発話数は 253 であり，そのうち 103 発話がシステムへの要求発話であった．

3. 提 案 手 法

本章では，本研究で用いたブースティングによるシス

テム要求と雑談の判別手法について述べる．本研究では，音声認識結果に対してブースティングを行うのではなく，音声認識の中にシステム要求検出を組み込む手法を用いた．これにより，認識の仮説を含む多くの情報をシステム要求検出に用いることが可能となる．また，システム要求検出の結果と音声認識の結果の整合性をとることも可能となる．

システム要求か否かを $s \in (request, chat)$，音声認識単語列を $\mathbf{W} = (w_1, \cdots, w_n)$，観測音声特徴系列を $\mathbf{O} = (o_1, \cdots, o_t)$ とすると，音声認識と統合されたシステム要求検出は以下のように定式化される．

$$
\begin{aligned}
(\hat{s}, \hat{\mathbf{W}}) &= \underset{(s,\mathbf{W})}{\operatorname{argmax}} P(s, \mathbf{W}|\mathbf{O}) \\
&= \underset{(s,\mathbf{W})}{\operatorname{argmax}} P(\mathbf{O})^{-1} P(s, \mathbf{W}, \mathbf{O})
\end{aligned}
$$

ここで，$P(s, \mathbf{W}, \mathbf{O})$ をベイズの定理により，以下の二通りに展開できる．

$$P(s, \mathbf{W}, \mathbf{O}) = P(s) \cdot P(\mathbf{W}|s) \cdot P(\mathbf{O}|\mathbf{W}, s) \quad (1)$$

$$P(s, \mathbf{W}, \mathbf{O}) = P(\mathbf{W}) \cdot P(\mathbf{O}|\mathbf{W}) \cdot P(s|\mathbf{W}, \mathbf{O}) \quad (2)$$

式 1 の定式化は，言語モデル・音響モデルが s に依存するようなモデルを用いる手法となる．ただし，高い識別性能は期待できないため，今回は利用しなかった．

式 2 の定式化は，言語モデル・音響モデルは通常のものを用いる．加えて，認識仮説 $\mathbf{W}$ や観測音声 $\mathbf{O}$ から直接 s を推定するような確率モデルが存在する．$\mathbf{W}$ や $\mathbf{O}$ から直接 s を推定するという意味において，このモデルは識別的なモデルである．識別的なモデルとしては，Support Vector Machines (SVM) やブースティングが考えられる．本研究では，従来から用いてきたブースティングを採用した．ただし，ブースティングは確率に基づく手法ではないため，音響モデル，及び言語モデルとの整合性をとるためには，ブースティングの出力結果を確率化する必要がある．本研究では，完全な確率とは言えないものの，ブースティングの出力スコアを sigmoid 関数を用いて疑似確率化して用いるものとした．sigmoid 関数は図 2 に示すような関数であり，最小値 0，最大値 1 となる．また，識別境界付近を詳細にモデル化できる特徴を持つ．ブースティングの出力スコア $f(\mathbf{W}, \mathbf{O})$ を sidmoid 関数で疑似確率化することにより，モデル $P(s|\mathbf{W}, \mathbf{O})$ は，

$$
\begin{aligned}
P(s = request|\mathbf{W}, \mathbf{O}) &= sigmoid(f(\mathbf{W}, \mathbf{O})) \\
P(s = chat|\mathbf{W}, \mathbf{O}) &= 1 - sigmoid(f(\mathbf{W}, \mathbf{O}))
\end{aligned}
$$

と定式化できる．ここで，w_1 及び w_0 は sigmoid 関数における重み係数であり学習により推定する．ただし，本研究では，$\mathbf{O}$ は用いず，$\mathbf{W}$ のみを用いた．次節で，ここで用いたブースティングアルゴリズムについて述べる．

3.1 ブースティング

本研究では，ブースティングの手法として AdaBoost を用いた．AdaBoost は，いくつもの識別器を組み合わせてひとつの高度な識別器を構成する *ensenble learning method* のひとつである．Schapire ら [4] が提案している学習のアルゴリズムを図 3 に示す．図中，I は，$I(true)$ ならば 1，$I(false)$ ならば -1 となる．ϵ_t が 0.5 未満の弱学習器を見つけ続けることができれば，学習誤差 0 の最終学習仮説を生成できる．また，未知のサンプルに対する汎化誤差も小さくできることが実験的に報告されている [5], [6]．一方，雑音を有するサンプルの場合，過学習を起こすことが報告されている．これに対しては，AdaBoost の学習過程をマージン最大化ととらえ，SVM における Soft Margins の概念を導入した手法も提案されている [7], [8]．本研究では，認識結果を扱うため，サンプルには多くの雑音が乗っているものと考えられる．このことから，通常の AdaBoost ではなく，Soft Margins 付きの AdaBoost を用いることとした．

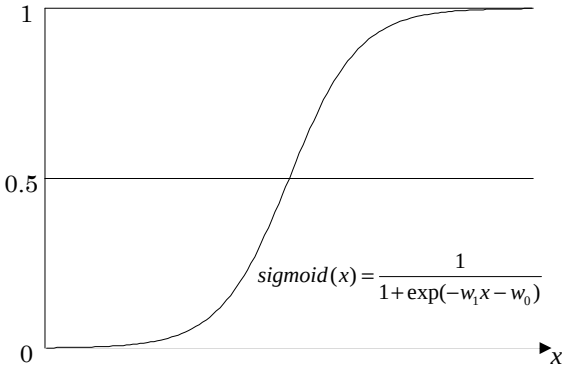

図 2　sigmoid 関数
Fig. 2　sigmoid function.

AdaBoost を用いたテキスト分類手法としては，文献 [4], [9] などが提案されている．これらの文献では，テキスト分類のための弱学習器として，Decision Stumps が用いられている．Decision Stumps とは，ある素性の有無に基づいて分類を行う単純な手法である．素性には，単語や単語 bigram，ラベル付き順序木などが用いられる．学習時には，学習サンプルを最もうまく分類するような “素性” を選択し，その際の重みを得る．識別時には，学習によって得られた全ての素性について，サンプル中にその素性があれば，クラス y に重み α の投票を行うということを繰り返し，最終的に重みの大きかったクラスと判別する．

3.2 ブースティングと音声認識の統合

前述の通り，ブースティングによる出力スコアを sigmoid 関数に当てはめることにより疑似確率化し，音声認識との統合を行う．弱識別器の数を T 個，弱識別器を $h_t(\mathbf{W})$，弱識別器の重みを α_t とすると，

$$P(s|\mathbf{W}, \mathbf{O}) = sigmoid(\sum_t \alpha_t h_t(\mathbf{W}))$$

Input: n examples $Z = \{(\mathbf{x}_1, y_1), \cdots, (\mathbf{x}_n, y_n)\}$
Initialize: $w_1(\mathbf{z}_i) = 1/n$ for all $i = 1 \cdots n$
Do for $t = 1, \cdots, T$,

（1） Train a base learner with respect to weighted example distribution w_t and obtain hypothesis $h_t : \mathbf{x} \mapsto \{-1, 1\}$

（2） Calculate the training error ϵ_t of h_t:

$$\epsilon_t = \sum_{i=1}^{n} w_t(\mathbf{z}_i) \frac{I(h_t(\mathbf{x}_i) \neq y_i) + 1}{2}.$$

（3） Set

$$\alpha_t = log \frac{\epsilon_t}{1 - \epsilon_t}$$

（4） Update example distribution w_t:

$$w_{t+1}(\mathbf{z}_i) = \frac{w_t(\mathbf{z}_i) \exp\{\alpha_t I(h_t(\mathbf{x}_i) \neq y_i)\}}{\sum_{j=1}^{n} w_t(\mathbf{z}_i) \exp\{\alpha_t I(h_t(\mathbf{x}_i) \neq y_i)\}}.$$

Output: final hypothesis:

$$f(\mathbf{x}) = \frac{1}{||\alpha||_1} \sum_{t=1}^{T} \alpha_t h_t(\mathbf{x}).$$

図 3　AdaBoost のアルゴリズム
Fig. 3　AdaBoost algorithm.

$$= \frac{1}{1 + \exp(-w_1 \sum_t \alpha_t h_t(\mathbf{W}) - w_0)}$$

となる（$s = request$ の場合）．sigmoid 関数のパラメータ w_1 及び w_0 は勾配法により学習する．ただし，学習をしすぎてしまうと sigmoid 関数の識別境界付近の勾配が急峻になりすぎてしまい，0 か 1 かの二値に近づいてしまうため，ある程度学習された時点で推定を止めるものとした．

ブースティングによるシステム要求検出と音声認識（音響モデル・言語モデル）を統合することにより，1-best の認識結果のみからシステム要求検出を行うのではなく，複数の認識仮説を利用できるようになる．また，その際，どの仮説からのシステム要求検出結果を用いればいいかについて，確率の枠組みの中で統一的に解を選択することが可能となる．例として，図 4 のような場合を考える．例はロボットタスクのものである．ここでは，「こっちに来て」「こっちにして」の 2 つの仮説が考えられる．次に，それぞれの仮説に対し，ブースティングによってスコアを求める．例として，「こっちに来て」は高いスコアでシステム要求，「こっちにして」は低いスコアで雑談と判定されるものとする．このスコアと音声認識の確率との統合を行う．ここで，ブースティングによる出力スコアは確率ではないため（値域も 0〜1 ではない），sigmoid 関数を用いて疑似確率化を行う．例では，「こっちに来て」は高確率でシステム要求，「こっちにして」は中確率で雑談，のようにブースティング出力スコアが（疑似）確率に変換される．統合後，最も確率の高い仮説が選択される．例えば，2 つの仮説の音声認識確率が同程度であれば，システム要求発話として自然な「こっちに来て」が

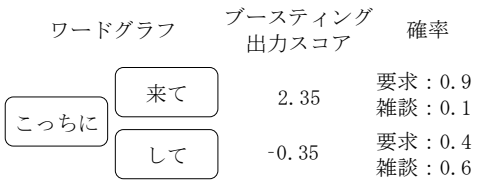

図 4　ブースティングによる出力スコアの確率化例
Fig. 4　An example of conversion of boosting score to probabilities.

選択される．逆に，2 つの仮説のうち，「こっちにして」の音声認識確率が顕著に高ければ，ブースティングによる出力スコアの違いを超えて「こっちにして」が選択される．システム要求検出と音声認識を確率の枠組みで統合することにより，認識仮説をシステム要求検出に利用することが可能となり，また，統一的な基準を基にして認識仮説，及びシステム要求か雑談かを判別することができる．

4. 実　　験

本節では，提案手法を用いたシステム要求検出実験について述べる．実験のタスクとして，2. 節で述べたものを用いた．ブースティングの学習は，音声認識器による 1-best の認識結果を学習データとして行った．ただし，検索キーワードはクラス化したものを学習に用いた．また，このとき学習されたブースティングのモデルと，同一の学習データから sigmoid 関数のパラメータの学習を行った．その後，提案手法による認識を行った．実験はすべて，10 folds のクロスバリデーションによって行った．評価は，システム要求検出の再現率・適合率・F 値によって行った．

比較手法として，システム要求発話のみで trigram を構築し，認識結果の単語信頼度の平均を閾値で区別する手法を用いた．

次節で，実験で用いた音声認識器について述べる．

4.1　音声認識条件と結果

ベースラインの音響モデルは，日本語話し言葉コーパス (CSJ: Corpus of Spontaneous Japanese) モニター版 [10] のうち，男性話者 200 名の講演音声を用いて作成した．音響分析条件と HMM の仕様を表 2 に示す．これらの条件で音響モデルを作成し，さらに，MLLR+MAP [11] により音響モデル適応を行った．音響モデル適応は，テストセットを含め，クローズドの適応を行った．これは適応のためのデータが少量しか得られなかったためである．

言語モデルは，実験で用いた発話を書き起こしたテキストから作成した．ただし，テストセットに対しオープンとなるように，テキストを 2 つに分割し，2 つの言語モデルを作成した．音声認識辞書については，2. 章で述べた通り，情報検索のためのインデックスを検索キーワードクラスとしてクラス化し，音声認識辞書に追加した．

表 2　音響分析条件と HMM の仕様
Table 2　Condition of acoustic analysis and HMM specification.

音響分析	サンプリング周波数	16kHz
	特徴パラメータ	MFCC(25 次元)
	フレーム長	20ms
	フレーム周期	10ms
	窓タイプ	ハミング窓
HMM	タイプ	244 音節
	混合数	32 混合
	母音 (V)	5 状態 3 ループ
	子音+母音 (CV)	7 状態 5 ループ

表 3　AdaBoost によって選択された素性語
Table 3　Selected features by AdaBoost.

システム要求	雑談
ニュース+</s> って+何	で な そろそろ とか
何+</s> たい+</s>	、+<KEYWORD>
<KEYWORD>+</s>	<KEYWORD>+って
リスト 見せ の+ニュース	これ やっ やん なんか
<s>+<KEYWORD> etc.	etc.

検索キーワードは全体で約 4000 語であった．

4.2　システム要求識別モデルの学習

ブースティングのモデルの学習は，書き起こしテキストを用いて，また音声認識結果から得られた 1-best の結果を用いて行った．素性には，unigram 及び bigram を用いた．このときに書き起こしテキストからの学習で選択された素性語の例を表 3 に示す．ロボットやカーナビタスクと同様，システム要求に投票を行う素性には bigram が多く選択されている．これは，システム要求発話がある程度決まったフレーズによって行われているためと考えられる．また，システム要求の素性には <s> や </s> との組み合わせが多数存在することから，システム要求発話は，ある程度会話を区切った上で為されているものと考えられる．ただし，従来と異なる点は，<KEYWORD> がシステム要求・雑談の両方に bigram 素性として現れている点である．<KEYWRDO> が文頭・文末に現れる場合はシステム要求の素性となるが，句点や「って」と結びついて現れる場合は雑談素性となっている．すなわち，<KEYWORD> 単体の有無にはシステム要求と雑談の識別能力はあまりなく，どのように用いられているか，という点が識別に寄与していると考えられる．

ここで得られたブースティングモデル，及び 1-best の認識結果から sigmoid 関数のパラメータを推定した．パラメータの推定は勾配法を用いて繰り返し計算で行う．ただし，繰り返しすぎると，sigmoid 関数の識別境界付近の傾きが急峻になりすぎてしまい，出力が 0 か 1 かの二値に近くなってしまう．そこで，パラメータの更新が緩やかになって来た時点，本研究では，パラメータの更新の比率が 0.01 以下となった時点で学習を止めるようにした．

4.3　システム要求検出

4.1 の条件で音声認識を行い，出力されたワードグラフからシステム要求検出実験を行った．また，システム要求のうち，

- NEWS: ニュース動画を検索
- LIST: トップニュースリストの表示
- NEXT: 次の映像の再生

のどの要求かの分類も行った．これは，NEWS かそれ以外（雑談+LIST+NEXT）のように one-vs-rest 方式で識別を行った．結果として，それぞれの手法において F 値が最も高かったケースを図 5 に示す．図中の REQUEST が，システム要求と雑談の判別結果である．また，凡例の意味はそれぞれ以下の通りである．

- 書き起こし：書き起こしテキストから学習したブースティングのモデルを用いて，書き起こしテキストを分類（認識誤りの影響のないクリーンテスト．システムの上限）
- ミスマッチ：書き起こしテキストから学習したブースティングのモデルを用いて，提案手法で分類
- 認識：認識結果の 1-best から学習したブースティングのモデルを用いて，提案手法で分類
- 信頼度：認識結果の信頼度の平均を閾値を用いて分類（比較手法）

実験の結果，システムの上限である，“書き起こし” から比べ，“ミスマッチ” では約 20%，“認識” では約 13%性能が低下した．どちらも認識誤りを原因とする性能の低下と考えられる．“ミスマッチ” と比べ，“認識” の性能低下が少ない理由として，“認識” では，音声認識を誤りやすい単語が，識別のための素性として選択されなかったことが考えられる．すなわち，認識誤りを含んだ上で，それでもコヒーレントな単語だけが素性として選択されたと考えられる．比較手法である “信頼度” は，“ミスマッチ” よりもわずかに低い結果となった．

システム要求の中での分類については，“NEWS” が，“LIST”，“NEXT” と比べて低い性能となった．これは，“LIST”，“NEXT” がロボットやカーナビタスクと同様，固定された単語から構成されるのに対し，“NEWS” は「○○のニュースを見せて」のように，○○の部分が変化するためと考えられる．認識誤りの内訳としては，「○○」のようにキーワードのみで検索を要求するような場合，キーワードの認識を雑談の素性語に誤って認識されている場合などが見られた．

5.　ま　と　め

本稿では，WWW 上でのニュース動画に対する検索タスクにおいて，システムへの要求と雑談の判別を行う手法について述べた．情報検索のためのインデックスをクラス化して音声認識辞書に登録することにより，個々の

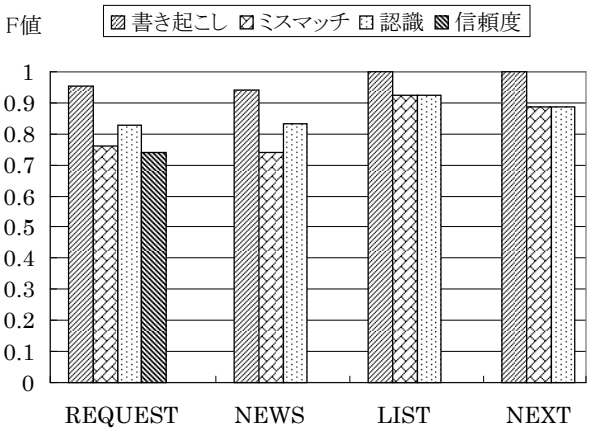

図 5　システム要求判別結果（信頼度はタスクの中身までの分類能力はないため REQUEST のみ）
Fig. 5　Result of system request discrimination.

検索キーワードではなく，キーワードクラスを素性としてブースティングを行った．これにより，情報検索のインデックスが変化した場合や，コーパス中では検索したことのない検索キーワードについてもブースティングモデルを再構築することなく，要求検出が可能である．また，システムへの要求発話に対し，さらに細かな分類を行った．コマンド中に検索キーワードを含むニュース検索要求では，精度低下が見られた．

今後の課題として，基礎的な音声認識性能の向上による性能改善，及び発話のコンテキスト情報の利用があげられる．

文　　献

[1] 田中克幸，滝口哲也，有木康雄，“ＮｅｔＴｖ：ＮｅｔＮｅｗｓとテレビ放送のクロスプラットホームにおける動画のインデキシングと音声検索，” 情報処理学会データベースシステム研究会研究報告，2007-DBS-141，pp.59-66，2007-01.

[2] 佐古淳，山形知行，滝口哲也，有木康雄，“音声認識との統合によるシステム要求検出，” 第９回音声言語シンポジウム，SP2007-120，pp.143-148，2007-12.

[3] 松本裕治，“形態素解析システム「茶筅」，” 情報処理，Vol.41，No.11，pp.1208-1214，2000.

[4] R.Schapire, Y.Freund, P.Bartlett, and W.Lee, “Boosting the margin: A new explanation for the effectiveness of voting mathods,” Annals of Statistics, vol.26, no.5, pp.1651-1686, Oct. 1998.

[5] Y.Freund and R.Schapire, “Experiments with a new Boosting algorithm,” Proc. 13th International Conference on Machine Learning Bari, Italy Morgan Kaufmann, pp.148-146, July 1996.

[6] H.Schwenk and Y.Bengio, “Adaboosting neural networks,” Proc. ICANN’97, vol.1327 of LNCS Berlin Springer, pp.967-972, Oct. 1997.

[7] G.Ratsch, T.Onoda, and K.-R. Muller, “Soft Margin for AdaBoost,” Machine Learning, vol.42, no.3, pp.287-320, March 2001.

[8] 小野田崇，“Boosting の過学習とその回避，” 電子情報通信学会論文誌，Vol.J85-D2，No.5，pp. 776-784，2002 年 5 月.

[9] 工藤拓／松本裕治，“半構造化テキストの分類のためのブースティングアルゴリズム，” 情報処理学会論文誌，Vol.45，NO.9，2004 年 9 月.

[10] 古井貞熙，前川喜久雄，伊佐原均，“『話し言葉工学』プロジェクトのこれまでの成果と展望，” 第２回話し言葉の科学と工学ワークショップ，pp.1-6，2002.

[11] 緒方淳，有木康雄，“音素事後確率に基づく信頼度を用いた音響モデルの教師なし適応，” 信学技報，SP2001-105，2001.