Vol.2012-SLP-92 No.7
2012/7/20

音声情報案内システムにおける Bag-of-Words を特徴量とした無効入力の棄却

真嶋 温佳[1,a)]　藤田 洋子[1]　トーレス ラファエル[1]　川波 弘道[1]　原 直[1]　松井 知子[2]　猿渡 洋[1]
鹿野 清宏[1]

概要：実環境音声情報案内システムでは，雑音等の非音声やユーザ同士の背景会話など，システムへの入力として不適切な入力が存在する．これらの入力はシステムの誤作動・誤認識の原因となるので，無効入力として棄却して応答処理を行わないことが重要である．一般に，有効入力と無効入力との識別には GMM (Gaussian Mixture Model) による方法など，音響的な情報に基づく方法が用いられることが多い．しかし，入力データに含まれる言語的な情報を使うことにより，システムのタスクも考慮した，より高精度な有効入力と無効入力の識別が可能になると考えられる．そこで本論文では，音声認識結果から得られる Bag-of-Words (BOW) を特徴量として，サポートベクターマシン（SVM）および最大エントロピー法を用いた無効入力の識別を検討した．実環境音声情報案内システム「たけまるくん」の入力データを用いた実験では，GMM を用いた SVM による無効入力の識別と較べ，F 尺度を 81.73%から 83.61%に改善することができた．また，BOW，GMM による音響尤度，発話時間，SNR を組み合わせた場合，F 尺度を 86.57%まで改善することができた．

キーワード：無効入力識別，Bag-of-Words, サポートベクターマシン，最大エントロピー法，音声情報案内システム

Invalid Input Rejection Using Bag-of-Words for Speech-Oriented Guidance System

Majima Haruka[1,a)]　Fujita Yoko[1]　Torres Rafael[1]　Kawanami Hiromichi[1]　Hara Sunao[1]
Matsui Tomoko[2]　Saruwatari Hiroshi[1]　Shikano Kiyohiro[1]

Abstract: On a real environment speech-oriented information guidance system, a valid and invalid input discrimination is important as invalid inputs such as noise, laugh, cough and utterances between users lead to unpredictable system responses. Generally, acoustic features are used for discrimination. Comparing acoustic likelihoods of GMMs (Gaussian Mixture Models) from speech data and noise data is one of the typical methods. In addition to that, using linguistic features is considered to improve discrimination accuracy as it reflects the task-domain of invalid inputs and meaningless recognition results from noise inputs. In this paper, we introduce Bag-of-Words (BOW) as a feature to discriminate between valid and invalid inputs. Support vector machine (SVM) and maximum entropy method (ME) are also employed to realize robust classification. We experimented the methods using real environment data obtained from the guidance system "Takemaru-kun." By applying BOW on SVM, the F-measure is improved to 83.61%, from 81.73% when using GMMs. In addition, experiments using features combining BOW with acoustic likelihoods from GMMs, Duration and SNR were conducted, improving the F-measure to 86.57%.

Keywords: invalid inputs discrimination, bag-of-words, support vector machine, maximum entropy method, speech-oriented guidance system

1. はじめに

音声認識システムを応用した音声情報案内システムは主

[1] 奈良先端科学技術大学院大学
Nara Institute of Science and Technology
[2] 統計数理研究所
The Institute of Statistical Mathematics
[a)] haruka-m@is.naist.jp

に，音声区間検出，音声認識，応答生成により構成され，これらの処理は入力音声に対して順次処理されていく．実環境における音声情報案内システムへの入力には，システムとして適切な発話（有効入力）以外の様々な入力が存在し，このような入力はシステムの誤認識や誤作動の原因となる．特に Push-to-talk などの機構を持たず，常に音声入力を受け付けるような音声情報案内システムでは，システムが応答すべきではない入力音声は非常に多い [1]．特に，応答生成処理はシステムが対応するドメインによって肥大化する可能性があり，すべての入力に対して応答処理をすることはシステム負荷の観点から避けるべきである．従って，システムにとって不適切な入力（無効入力）はできる限り応答生成処理に送られる前に棄却することが望ましい．

これまでにも，音声認識結果に対する統計的仮説検定による発話照合手法が研究されている [2]．この手法では，ある音声認識結果を含む音声発話区間が入力された場合にその発話を受理すべきであるという帰無仮説と棄却すべきであるという対立仮説を立てる．そして，それぞれの仮説を条件とした入力音声の条件付き確率分布を事前に学習することができれば，それらの尤度比を検定統計量とした仮説検定を行うことができる．つまり，未知の入力発話に対して帰無仮説が棄却された場合，その入力発話は高い確率で受理すべきではないと判定される．この手法は，音声の確率分布を考えるということで一種の生成的なアプローチと考えられるが，より直接的な識別的なアプローチも考えられる [3]．また，Lane らは音声対話システムへのドメイン外発話の検出手法を提案している [4]．彼らは，N-best 認識仮説から得られた Bag-of-Words（BOW）を特徴量として発話のトピック推定を行い，そのトピック推定結果を用いてドメイン外発話の検出を行っている．これらの手法は，特徴量として音声認識結果を利用しており，認識結果の後処理として位置づけられる．

一方で，音声認識の前処理として音響的特徴量を用いた手法も検討されており，たとえば，メル周波数ケプストラム係数（Mel-Frequency Cepstral Coefficient: MFCC）特徴量に対する混合ガウス分布モデル（Gaussian Mixture Model: GMM）から算出される音響尤度による最尤判別に基づく音声と雑音の識別 [5], [6], [11] などがあげられる．また，音声区間検出においても，音声認識結果の言語的制約を用いる手法が提案されており [7]，これは言語的な情報も音声と雑音の識別に有効であることを示唆している．

以上を踏まえて，本研究では音声認識の後処理で用いられる音声認識結果に基づく言語的特徴量と，音声認識の前処理で用いられる音響的特徴量の併用による，音声情報案内システムへの無効入力の検出手法を提案する．本論文では音声認識結果から得られた BOW を主な特徴量として扱う．音声認識結果に含まれる有効入力に出現しやすい単語や無効入力の認識結果の傾向を利用することができれば，システムのタスクを考えた上で有効入力と無効入力を識別することが可能になると考える．また，BOW に音響尤度などを組合せることで識別精度の向上も行う．識別手法として，サポートベクターマシン（Support Vector Machine: SVM）[8] および最大エントロピー法（Maximum Entropy method: ME）[9], [10] を用いる．

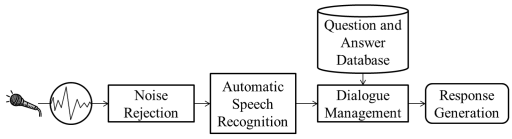

図 1 「たけまるくん」[1] における応答処理の流れ
Fig. 1 Processing flow in “Takemaru-kun”

本論文の構成は以下の通りである．第 2 章では実験に用いた音声情報案内システム「たけまるくん」とこの「たけまるくん」によって収集された音声コーパスについて述べる．第 3 章では本論文で扱う特徴量及び SVM や ME を用いた無効入力の識別手法を述べる．第 4 章では本論文の提案する特徴量と識別手法を用いた識別実験の結果を示し，第 5 章で結論を述べる．

2. 音声情報案内システム「たけまるくん」

2.1 システムの概要

「たけまるくん」[1] は，生駒市北コミュニティセンター内に設置された音声情報案内システムである．2002 年 11 月より運用を開始し，現在までの約 10 年にわたり運用を継続している．「たけまるくん」に対してユーザが発話によって質問すると，合成音声とアニメーションを用いて，エージェントが応答する．「たけまるくん」の主な応答内容は，コミュニティセンター内の施設案内，周辺の観光案内，エージェント（たけまるくん）自身に対する質問への応答，現在時刻・天気・ニュースなどである．これらとは別に，ユーザが発話した単語を Web で検索する「Web 検索モード」も利用できる．「たけまるくん」は，音声対話システムの実環境対話データ収集及び対話システムのフィールドテストを兼ねて運用されている．本システムの構成を**図 1** に示す．「たけまるくん」は，音声のみを入力インタフェースとしている．そのため，ユーザはシステムに対して話しかけるという単純な行動のみでシステムから情報を得ることができ，自由度の高いシステムとなっている．しかし，無効入力によってシステムが誤作動すると，ユーザにとって非常に扱いにくいシステムとなる．図 1 の “Noise Rejection” では，Lee ら [11] による 5 クラスの GMM を用いた音響尤度に基づく棄却が行われており，単純に有効と無効の 2 クラスのモデルを作るよりも雑音それぞれのクラスのモデルを作成することが有効であると示されている．しかし，本論文で無効入力としている「無効発話」は実験データには含まれていない．したがって，「無効発話」も含めた無効入力の識別は重要な課題である．

2.2 収集データ

「たけまるくん」は 2002 年 11 月より運用を開始し，以

表 1 「たけまるくん」の入力データの分類結果 (2002 年 11 月から 2004 年 10 月まで)

Table 1 Classification result on input data of "Takemaru-kun" (from Nov. 2002 till Oct. 2004)

カテゴリ			発話数	合計
有効入力	大人発話		20436	
	子供発話		85889	106325
無効入力	無効発話	背景会話	26319	122939
		発話不明瞭	13348	
		意味のない発話	11991	
		音声区間検出ミス	12937	
		オーバーフロー	1417	
		レベル不足	7347	
	咳		727	
	笑い声		6232	
	雑音		50756	

降現在までのすべての入力データを収録している．この内，最初の 2 年間分のデータは聴取による書き起しと，有効入力または無効入力のラベル，年齢層，性別，雑音などのタグが付与され，データベースとして整備が進められている．本論文における有効入力，無効入力の分類はこのラベルに従ったものである．詳細な分類を**表 1** に示す．表 1 において，無効入力は，人の音声による「無効発話」,「咳」,「笑い声」，及びそれ以外の非音声の「雑音」に大きく分類される．「無効発話」にはさらに詳細なタグが付けられている．「背景会話」とは発話者の背後で他人の会話が重なって聞こえるもの及び明らかにシステムに対する発話ではなくマイクの周辺で会話されている発話，「発話不明瞭」とは音声が聞き取りづらく客観的に判別できないもの，「意味のない発話」とはフィラーや「マイクテスト」などのシステムからの情報取得が目的ではない発話，「音声区間検出ミス」とは文頭もしくは文末が欠損している発話，「オーバーフロー」は発話者の声が大きすぎて音割れを起こしている発話，「レベル不足」は入力音が小さすぎて発話内容が聴取できない発話を指す．なお，これらのタグは重複を許している．また，雑音タグが与えられていても，発話内容がシステムの入力として有効である場合は有効入力として分類しており，表 1 の無効入力には集計していない．

3. BOW 特徴量を用いた無効入力の識別

3.1 特徴量

特徴量の選択は，識別性能に大きな影響を与える．本論文では以下の 4 種類の特徴量を検討した．

- **GMM による音響尤度（GMM）**
  入力音の 6 クラスの各 GMM に対する尤度の時間平均を要素とする 6 次元のベクトル．GMM は音響特徴量が大きく異なると考えられるクラスごとに作成した．有効入力として「大人発話」,「子供発話」，無効入力として「無効発話」,「咳」,「笑い声」,「雑音」の計 6 クラスの GMM を用いた．GMM 作成に使用したデータは「たけまるくん」によって実環境で収集されたデータである．
- **Bag-of-Words (BOW)**
  音声認識結果の N-best に含まれている単語の出現頻度を要素とした特徴量ベクトル．数え上げる単語は，学習データの音声認識結果から作られた単語辞書中にあるものに限る．
- **発話時間（Duration）**
  音声認識エンジン Julius[14] による振幅と零交差法に基づく音声区間検出により，一発話とみなされた入力音の時間長．
- **信号対雑音比（Signal to Noise Ratio: SNR）**
  一発話ごとに算出した SNR の値．入力音をフレームに分割し，便宜的にそのフレームの中で平均パワーの大きいフレームの上位 10%を信号区間，平均パワーの小さいフレーム下位 10%を雑音区間と考え，次式により求める．

$$\mathrm{SNR} = 10 \log_{10} \frac{P_S - P_N}{P_N}$$

  ここで，P_S は信号区間の平均パワー，P_N は雑音区間の平均パワーを表す．

3.2 識別手法

3.2.1 SVM による識別手法

SVM[8] は教師あり学習機械であり，2 クラス分類問題を対象とする．SVM は，与えられたデータを，カーネル関数によって高次元へと写像し，写像した空間において 2 クラスに分類する．その際に 2 クラス間のマージンが最大となる識別境界を求める．今，n 次元の特徴量ベクトル $\mathbf{x}_i \in R^n, i = 1, ..., l$ （l ：サンプル数）とラベル $y_i \in \{+1, -1\}$ のペア集合が与えられたとすると，次式にしたがって，2 クラス分類のための識別境界が求められる．ここで $\mathbf{w}$，b は識別関数のパラメータ，C はコストパラメータ，$\phi(\mathbf{x}_i)$ 特徴量ベクトルを高次元の空間へ写像する関数であり、この関数により特徴量ベクトルの非線形な分類ができるようになる．

$$\min_{\mathbf{w},b,\xi} \frac{1}{2} \quad \mathbf{w}^T\mathbf{w} + C\sum_{i=1}^{l} \xi_i$$
$$\text{subject to } y_i(\mathbf{w}^T\phi(\mathbf{x}_i) + b) \geq 1 - \xi_i,$$
$$\xi_i \geq 0, i = 1, ..., l.$$

ここで，ξ_i はスラック変数であり，これによりある程度の誤分類を許容しつつマージンの最大化を行う．また，この式を次式へと拡張することにより，正例 $(y_i = +1)$ と負例 $(y_i = -1)$ の数がアンバランスな問題に対処できる．

$$\min_{\mathbf{w},\mathbf{b},\xi} \frac{1}{2} \quad \mathbf{w}^T\mathbf{w} + C_+ \sum_{\{i:y_i=+1\}} \xi_i + C_- \sum_{\{i:y_i=-1\}} \xi_i$$

$$\text{subject to } y_i(\mathbf{w}^T\phi(\mathbf{x}_i)+b) \geq 1-\xi_i,$$
$$\xi_i \geq 0, i = 1, ..., l.$$

具体的には，分類誤りに対するコストパラメータ C を，正例を負例とする誤りのコストパラメータ C_+ と負例を正例とする誤りのコストパラメータ C_- に分けて，2 つの誤り率のバランスをとる．本論文では，いくつかのコストパラメータ C によって評価データの識別性能を求めて，最適値を求めた．C_+ と C_- は次式により与えられる．

$$C = C_+ + C_-$$
$$C_+ = \frac{N_-}{N_+ + N_-} \times C, \quad C_- = \frac{N_+}{N_+ + N_-} \times C$$

ここで，N_+ は正例のデータ数，N_- は負例のデータ数である．

一般に，SVM ではすべての特徴量ベクトルから各データのカーネル値を計算し，識別境界を求める．しかし，傾向の異なる複数の特徴量を用いる場合，各特徴量間の値の大きさの違いや次元数の違いを考慮する必要がある．そこで本論文では，マルチカーネル法 [12] を用いる．

データ x_i, x_j のカーネル値 $k(x_i, x_j)$ を i, j 成分とする行列（グラム行列）[13] を特徴量ごとに算出し，足し合わせてから識別境界を求める手法をマルチカーネル法と呼ぶ．特徴量の種類が M 個の時，マルチカーネル法は以下の式により表すことができる．

$$K(x_i, x_j) = \sum_{c=1}^{M} a_c k_c(x_i, x_j), \quad \sum_{c=1}^{M} a_c = 1, \ a_c \geq 0$$

ただし，$k_c(x_i, x_j)$ は $\phi(x_i)$ と $\phi(x_j)$ の内積で表されるカーネル関数である．

3.2.2 ME による識別手法

ME[9], [10] は，分類問題によく用いられる一般的な機械学習手法である．ME では，特徴量はモデルの制約に対応しており，複数の特徴量を統合することができる．クラスのラベル集合 E と特徴量の集合 D について，学習データのセット (E, D) が与えられたとき，以下の対数尤度を最大化することにより，$\Lambda = \{\lambda_i; i = 1, ..., l\}$ を学習する．

$$\log P(E|D, \Lambda) = \sum_{(e,d)\in(E,D)} \log \frac{\exp \sum_i \lambda_i f_i(e, d)}{\sum_{e'} \exp \sum_i \lambda_i f_i(e', d)}$$

ここで，f_i は特徴量に対する素性関数であり，λ_i は各素性関数に対する重みである．そして，学習された Λ を用いて，事後確率が最大となるクラスにデータを識別することができる．なお，ME で複数の特徴量を用いる場合，単純に特徴量の集合 D に新たな特徴量を追加する．

4. 有効入力と無効入力の識別実験

本実験の目的は，以下の二つである．第 3 章で挙げた特徴量を用いた SVM 及び ME による有効入力と無効入力の識別精度を評価すること及び，複数の特徴量を組み合わせて識別精度を評価することである．

表 2　実験データ
Table 2　Experiment data

	有効入力（正例）	無効入力（負例）	計
学習データ	7607	7274	14881
テストデータ	3782	3902	7684

表 3　GMM の学習に用いたデータ数
Table 3　Training data used in GMM training

有効入力	大人発話	1053
	子供発話	6554
無効入力	無効発話	3640
	咳	29
	笑い声	287
	雑音	3318

4.1 実験条件

本実験において使用するデータを**表 2** に示す．この学習データ，テストデータはそれぞれ「たけまるくん」によって得られた 1 ヶ月分の入力データである．なお，表 2 の学習データと GMM の学習データは一致させている．GMM の学習において，標本化/量子化は 16 kHz/16 bit，分析窓長は 25 msec，窓シフト長は 10 msec とし，MFCC (12 次元)，Δ MFCC，Δパワーを用いた．混合数は 128 とし，学習に用いたデータは**表 3** のとおりである．実験条件を**表 4** に示す．

本実験における評価尺度には，F 尺度 [19] を用いる．F 尺度は，適合率（P）と再現率（R）という，正確性と網羅性の総合的な評価尺度であり，以下の式で定義される．

$$\mathrm{F} = \frac{2 \cdot P \cdot R}{P + R}$$

ただし，

$$P = \frac{N_{\mathrm{TP}}}{N_{\mathrm{TP}} + N_{\mathrm{FP}}}, \quad R = \frac{N_{\mathrm{TP}}}{N_{\mathrm{TP}} + N_{\mathrm{FN}}}$$

であり，N_{TP} は有効入力を有効入力と識別した数，N_{FP} は無効入力を有効入力と識別した数，N_{FN} は有効入力を無効入力と識別した数である．

4.2 実験 1：各特徴量の識別性能の評価実験

第 3 章で述べた特徴量を個別に用いて，SVM 及び ME によって有効入力と無効入力の識別を試みる．第 3 章で述べた 6 クラスの GMM を用いた SVM による無効入力の識別手法を従来手法として考え，その他の特徴量を用いた時の SVM 及び ME による識別手法の結果と比較する．

実験 1 の結果を**図 2** に示す．BOW を特徴量とした場合，SVM，ME のどちらも従来手法より F 尺度が向上し

表 4　実験条件
Table 4　Experimental condition

音声認識	エンジン	Julius 4.2[14]
	言語モデル	「たけまるくん」の 2 年間分の書き起こし文から作ったモデル [7]
	音響モデル	JNAS[15] モデルを「たけまるくん」データで適応したモデル
	出力	10-best
形態素解析器		Chasen 2.3.3[16]
単語辞書サイズ		4488
SVM	SVM ツール	LIBSVM[17]
	カーネル関数	Radial Basis Function (RBF)
	パラメータ C	$10^{-2}, 10^{-1}, ..., 10^{4}$（10 倍刻み）
ME	ME ツール	Stanford Classifier 2.1.3[18]

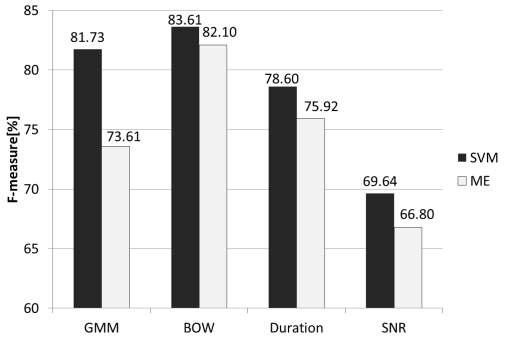

図 2　単一特徴量による無効入力の識別性能
Fig. 2　Result of invalid input discrimination using a single feature

ている．BOW を特徴量とした SVM が最良の結果であり，従来手法と比較し，F 尺度が 81.73%から 83.61%に改善した．また，BOW を特徴量とした ME でも，82.10%に改善したこのことから，無効入力の識別における BOW の有効性が示された．

4.3　実験 2：複数特徴量の組合せによる識別性能の評価実験

複数の特徴量を組み合わせて無効入力の識別を試みる．本実験では，特徴量を加えるごとに，F 尺度がどのように変化していくかを観察するため，実験 1 において有効であった特徴量を順番に加えていった．

実験 2 の結果を**図 3** に示す．最良の結果を示したのは，「BOW，GMM，Duration，SNR」のすべての特徴量を用いた SVM による識別手法であり，F 尺度は 86.57%に改善された．これは，従来手法と比較すると 4.84 ポイント改善できている．このことから，複数特徴量の組合わせは，無効入力の識別において有効であることが示された．

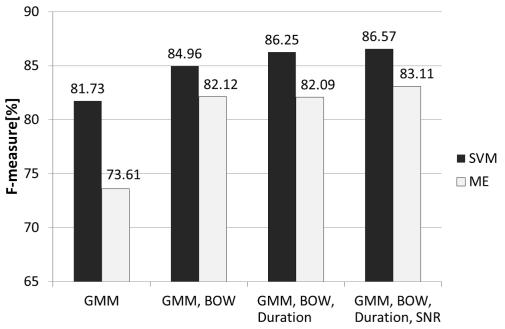

図 3　複数特徴量による無効入力の識別性能
Fig. 3　Result of invalid input discrimination using multiple features

4.4　考察

カテゴリ別の識別誤りの傾向を調査するため，カテゴリ別の識別誤り率（Error Rate: ER）を算出し，**表 5** に示す．ER は以下の式で定義される．

$$ER = \frac{N_e}{N_d} \times 100$$

ただし，N_d はテストデータ中のそのカテゴリに含まれるデータの数，N_e は N_d のうち誤ってカテゴリ外に識別されたデータの数である．GMM の音響尤度を用いた SVM による従来手法では「音声区間検出ミス」の ER が 80%近くあったが，BOW を用いた識別では，SVM 及び ME ともに 30%台に改善できている．「意味のない発話」および「発話不明瞭」においても，ER を 25%程度から 10%以下および 45%程度から 30%以下に改善できている．これは，BOW を用いることで入力データの言語的情報を識別結果に反映できたためと考えられる．また，BOW を用いることで，「有効入力」，「オーバーフロー」（ME のみ），「咳」，「笑い声」の ER が増えているが，これは音声認識誤りによるものだと考えられる．「オーバーフロー」は，どの手法でも ER が 40%を超えており，識別が困難なカテゴリであることが分かる．なお，従来手法と識別性能が最良の「BOW，GMM，Duration，SNR」を特徴量とした SVM を比較すると，ER が悪化しているのは「オーバーフロー」のみであり，それ以外のカテゴリでは全て改善された．

5.　まとめ

無効入力の識別手法として，BOW を特徴量とした SVM 及び ME による識別を提案した．GMM を用いた SVM による従来手法と比較して，F 尺度は，BOW を用いた SVM では 81.73%から 83.61%に，ME では 83.11%に改善できた．さらに，複数の特徴量を組み合わせることによって，SVM では F 尺度を 86.57%まで改善できた．以上の実験により，BOW 特徴量は無効入力の識別に有効であることが示されたが，いくつかの課題が残されている．BOW の次元数は対話ドメインに依存するため，別ドメインの対話コーパスに適用した場合の比較実験により，本手法の有効性を示す必要がある．また，本論文で用いた「たけまる

表 5　カテゴリ別の識別誤り率
Table 5　Discrimination error rate of each category

カテゴリ			テストデータに含まれる数	Error Rate				
				GMM (従来手法)	BOW		BOW, GMM, Duration, SNR	
				SVM	SVM	ME	SVM	ME
有効入力			3782	10.05	18.11	18.77	9.94	18.06
無効入力	無効発話	背景会話	929	19.59	15.17	18.51	9.58	13.89
		発話不明瞭	328	44.51	23.17	29.88	29.88	30.18
		意味のない言葉	728	24.31	3.85	7.42	11.13	8.93
		音声区間検出ミス	563	77.62	33.21	38.72	47.78	36.59
		オーバーフロー	62	41.94	40.32	45.16	53.23	53.23
		レベル不足	215	7.9	0	0	0	0
	咳		29	0	3.45	3.45	0	3.45
	笑い声		160	5.63	6.25	12.50	5.00	8.13
	雑音		1366	8.57	3.00	4.32	4.83	4.03

くん」データベースには大量のラベルなしデータがあるため，それらを活用した半教師あり学習（semi-supervised learning）[20] による識別性能向上が考えられる．

謝辞　本研究の一部は，戦略的創造研究推進事業「共生社会に向けた人間調和型情報技術の構築」(JST/CREST) の援助を受けて行われた．

参考文献

[1] R. Nisimura, A. Lee, H. Saruwatari, K. Shikano:Public speech-oriented guidance system with adult and child discrimination capability, *in Proc. ICASSP*, pp. 433–436 (2004).
[2] R.A. Sukka, C.-H. Lee: Vocabulary independent discriminative utterance verification for nonkeyword rejection in subword based speech recognition, *IEEE Trans. on Speech and Audio Processing*, Vol.4, No.6, pp. 420–429 (1996).
[3] T. Matsui, F.K. Soong, B.-H. Juang: Verification of multiple class recognition: a classification approach, *IEICE Trans. on Information and Systems*, Vol.E88-D, No.3, pp. 455–462 (2005).
[4] I.R. Lane, T. Kawahara, T. Matsui, S. Nakamura: Out-of-domain utterance detection using classification confidences of multiple topics, *IEEE Trans. on Acoustis, Speech, and Language Processing*, Vol.15, No.1, pp. 150–161 (2007).
[5] 中村敬介，西村竜一，李晃伸，猿渡洋，鹿野清宏：実環境音声情報案内システムにおける環境雑音および不要発話の識別，電子情報通信学会技術研究報告 SP2003-172，pp. 13–18 (2004).
[6] 鈴木智詞，竹内義則，松本哲也，工藤博章，大西昇：聴覚障害者のための警告音の識別，電子情報通信学会技術研究報告 Vol. 104，No. 695, SP2004-154-163, pp. 13–18 (2005).
[7] H. Sakai, T. Cincarek, H. Kawanami, H. Saruwatari, K. Shikano, A. Lee: Voice activity detection applied to hands-free spoken dialogue robot based on decoding using acoustic and language model, *Proceedings of the 1st international conference on Robot communication and coordination (ROBOCOMM2007)*, Article No. 16, 8 pages (2007).
[8] V.N. Vapnik: *The Nature of Statistical Learning Theory*, Springer (1995).
[9] A.L. Berger, S.D. Pietra, V.D. Pietra: A maximum entropy approach to natural language processing, *Computational Linguistics*, Vol.22, No.1, pp. 39–71 (1996).
[10] C. Manning, D. Klein: Optimization, maxent models, and conditional estimation without magic, *Tutorial at HLT-NAACL and ACL*, (2003).
[11] A. Lee et al.: Noise robust real world spoken dialog system using GMM based rejection of unintended inputs,*in Proc. ICSLP*, pp. 173–176 (2004).
[12] G. R. G. Lanckriet, N. Cristianini, P. Bartlett, L. E. Ghaoui, M. I. Jordan: Learning the kernel matrix with semidefinite programming, *Journal of Machine Learning Research 5*, pp. 27–72 (2004).
[13] C. M. Bishop: *Pattern Recognition and Machine Learning*, chapter 6, pp. 291–294 (2006).
[14] A. Lee, T. Kawahara, K. Shikano: Julius - an open source real-time large vocabulary recognition engine, *in Proc. Eurospeech*, pp. 1691–1694 (2001).
[15] K. Itou, M. Yamamoto, K.Takeda, T. Takezawa, T. Matsuoka, K. Shikano, T.Kobayashi, S.Itahashi, The design of the newspaper-based Japanese large vocabulary continuous speech recognition corpus, *in Proc. ICSLP*, vol. 7, pp. 3261–3264 (1998).
[16] 形態素解析器 Chasen，入手先 ⟨http://chasen-legacy.sourceforge.jp/⟩（参照 2012-05-31）.
[17] C. Chang, C. Lin: *LIBSVM: a library for support vector machines* , available from ⟨http://www.csie.ntu.edu.tw/˜cjlin/libsvm⟩ (accessed 2012-05-31).
[18] Stanford Classifier, available from ⟨http://nlp.stanford.edu/software/classifier.shtml⟩ (accessed 2012-05-31).
[19] C. D. Manning et al.: *Introduction to Information Retrieval*, chapter 8, pp. 154–157 (2008).
[20] X. Zhu: Semi-supervised learning literature survey, Computer Sciences, University of Wisconsin-Madison, No.1530 (2005).