音声言語情報処理 45−16
（2003．2．8）

機械学習を用いた複数の大語彙連続音声認識モデルの出力の混合

小玉　康広† 渡邊　友裕† 宇津呂武仁‡ 西崎　博光† 中川　聖一†

† 豊橋技術科学大学 工学部 情報工学系
‡ 京都大学大学院 情報学研究科 知能情報学専攻
†{kodama,watanabe,nisizaki,nakagawa}@slp.ics.tut.ac.jp, ‡utsuro@i.kyoto-u.ac.jp

あらまし 本論文では，機械学習を用いて複数の大語彙連続音声認識モデルの出力を混合するタスクに対して，SVM(Support Vector Machines) を適用する．SVM により複数モデルの出力を混合する規則を学習し，この規則を用いて，デコーダ，音響モデルの異なる 26 種類の大語彙日本語連続音声認識モデルの出力の混合を行ったところ，認識率最大の単独モデル，および，決定リスト学習や (重み付き) 多数決を用いた混合の単語認識率を上回る性能が達成できた．その単語誤り改善率は，認識率最大の単独モデルに対して最大で 72%，また，多数決法による複数モデル混合に対して最大で 36%という高い性能であった．

Combining Outputs of Multiple LVCSR Models by Machine Learning

Yasuhiro KODAMA†, Tomohiro WATANABE†, Takehito UTSURO‡,
Hiromitsu NISHIZAKI†, and Seiichi NAKAGAWA†

† Dpt. Information and Computer Sciences, Toyohashi University of Technology
‡ Dpt. Intelligence Sci. and Tech., Graduate School of Informatics, Kyoto University
†{kodama,watanabe,nisizaki,nakagawa}@slp.ics.tut.ac.jp, ‡utsuro@i.kyoto-u.ac.jp

Abstract We apply SVM learning technique to the task of combining outputs of multiple LVCSR models, where, as features of SVM learning, information such as the pairs of the models which output the hypothesized word are useful for improving the word recognition rate. Experimental results show that the combination results achieve a relative word error reduction of up to 72 % against the best performing single model and that of up to 36 % against ROVER.

1. はじめに

近年，音声認識結果の正解部分と誤り部分を分離することを目的として信頼度 (Confidence Measure) の研究が行なわれている (例えば，連続音声認識では [5], [11] など)．これまで提案されてきた信頼度尺度の多くは，いずれも，単一の認識エンジン・認識モデルが出力する認識結果を用いて，その正解部分と誤り部分を分離するというものであった．一方，連続音声認識の認識率そのものの向上を目的とする研究においては，複数の認識システムの出力を統合する方式 (ROVER 法 [2]) も提案され，一定の効果が報告されている．我々は，ROVER 法のような (重み付き) 多数決法が認識率の改善に効果的であることを考慮して，音声認識結果の正解部分と誤り部分を分離するための信頼度尺度として，複数の音声認識システムの出力の共通部分を用いる方法を提案し，その有効性を示した [9]．評価実験の結果では，デコーダおよび音響モデルが異なる二つのモデルについて，出力の共通部分の信頼度を評価したところ，最も高い性能が達成された．新聞読み上げ音声では，94%程度の単語正解率を 7%程度犠牲にすることにより，99%近い適合率を達成し，ニュース音声でも，72%程度の単語正解率を約 9%弱程度犠牲にすることにより，95%近い適合率を達成した．また，同一のデコーダを用いた場合は，新聞読み上げ音声で 98%程度，ニュース音声で 93〜94%程度という適合率であり，デコーダが異なる場合の性能を約 1%程度下回るものの，ほぼそれに匹敵する性能を達成した．

さらに，デコーダ，音響モデルの異なる 26 種類の大語彙日本語連続音声認識モデルについて，あらゆる可能な二つのモデル組の出力の間の共通部分の再現率・適合率を評価し，単語の品詞ごとあるいは音節数ごとに，最大の適合率を示すモデルの組合せが異なることを示した [7], [8]．また，機械学習 (決定リスト学習 [12]) の枠組を用いて，単語の品詞ごとあるいは音節数ごとに，最も適合率の高いモデルの組合せを選択的に組み合わせる規則を学習し，この混合規則を適用することにより，単独モデルの適合率・再現率を改善でき，二つのモデル組の出力の共通部分の適合率・再現率のうち，特に再現率が大幅に改善できることを示した [8]．さらに，ROVER 法のような (重み付き) 多数決法との比較においても，機械学習を用いた混合法の性能が上回ることを示した．

これまでの結果をふまえて，本論文では，機械学習を用

いて複数の大語彙連続音声認識モデルの出力を混合するタスクに対して，機械学習の各種手法の中でも，一般により性能が高いとされている SVM(Support Vector Machines) [10] を適用する．SVM により複数モデルの出力を混合する規則を学習し，この規則を用いて，デコーダ，音響モデルの異なる 26 種類の大語彙日本語連続音声認識モデルの出力の混合を行ったところ，決定リスト学習を用いた混合の単語認識率を上回る性能が達成できた．また，全 26 種類のモデルのうち単語認識率の上位 n $(3 \leqq n \leqq 26)$ 個のモデルの出力の混合を行ったところ，決定リスト学習や (重み付き) 多数決法では，混合結果の単語認識率が最大となったのは，26 種類のモデル全てを用いて混合を行った場合ではなく，単語認識率が上位のモデル十数種類に限定して混合を行った場合であった．そして，これらの十数種類のモデルに，単語認識率が下位のモデルを追加して混合を行った結果では，単語認識率の低下が観測された．この結果から，決定リスト学習や (重み付き) 多数決法による混合においては，性能の低いモデルが混入することにより，混合結果の単語認識率が低下する傾向があることが分かった．一方，SVM を用いた混合でも，単語認識率が上位のモデル十数種類の出力の混合を行った結果，混合結果の単語認識率が最大値に達したが，これらの十数種類のモデルに，単語認識率が下位のモデルを追加して混合を行っても大幅な単語認識率の低下は観測されなかった．このことから，SVM を用いた混合では，性能の低いモデルが混入しても，学習の段階でこれに対処することが実現できており，学習性能の高さを実証することができた．

2. 実験条件

2.1 大語彙日本語連続音声認識モデル

大語彙連続音声認識モデルとしては，SPOJUS [1] (音響モデル [6] は，12kHz/16kHz サンプリング，フレーム周期 8/10ms，特徴ベクトルはセグメント単位/フレーム単位の MFCC の二種類，音節モデル, 無音モデル有・無二通り，全/対角共分散行列，継続時間制御/自己遷移ループ，等合計 18 種類) および Julius [4] (音響モデルは，16KHz サンプリング，フレーム周期 10ms，特徴ベクトルはフレーム単位の MFCC，トライフォン/モノフォン/PTM/音節モデル，無音モデル有・無二通り，の合計 8 種類) を使用した．言語モデルは，毎日新聞 (45ヵ月分) または NHK 汎用ニュース原稿 (5 年分) から作成した tri-gram モデル (語彙数 2 万) を用いた．言語モデルの各語彙は，単語・品詞・読みの三項組で表現されており，認識結果についても，単語・品詞・読みの三項組の列として出力される．

2.2 評価データ

評価データとしては新聞読み上げ音声コーパス (JNAS) [3] の 100 文 (男性話者 10 人,1565 単語) の音声，および，NHK のニュース「ニュース 7」と「おはよう日本」(1996 年 6 月 1 日) の 175 文 (男性話者,6813 単語) の二種類を使用した．新聞読み上げ音声の認識精度は，SPOJUS で単語正解率 90.2∼78.1%，単語正解精度 85.3∼51.0%，Julius で単語正解率 93.0∼72.7%，単語正解精度 90.4∼69.4%，ニュース音声の認識精度は，SPOJUS で単語正解率 70.7∼55.4%，単語正解精度 62.8∼ 36.2%，Julius で単語正解率 71.7∼49.0%，単語正解精度 68.8∼ 39.7%であった

3. 機械学習による複数の大語彙連続音声認識モデルの出力の混合

各単語を出力したモデルの情報・単語の品詞・音節数・音響スコア・言語スコアを素性として，機械学習により，複数の大語彙連続音声認識モデルの出力を混合する規則の学習を行い，この規則を適用することにより複数モデルの出力の混合を行う．まず，2.2 節の評価音声データ (新聞読み上げ音声 100 文，あるいは，ニュース音声 175 文) を，話者が重複しないように，訓練データセットと評価データセットに分割する．そして，訓練データセットを用いて複数モデルの出力を混合する規則を学習し，この混合規則を評価データセットに適用することにより，複数モデルの出力の混合の評価を行う．混合規則の学習・適用に際しては，まず，混合の対象となる複数のモデルの出力の単語列に対して DP マッチングを行い，単語ラティスを構成する．そして，単語ラティス中の各単語の正誤を，機械学習の際に判別対象とするクラス c とし，このクラスを決定するための規則を学習する．

3.1 SVM

SVM [10] は，個々の素性を次元とする多次元空間中の点として記述された各事例に対して，全事例を二クラスに分類する境界面を学習することによりクラス判別規則を学習するという，機械学習の一つの手法である．SVM は，境界面をはさんで最も近い位置にある二つの事例の間の距離 (マージン) を最大化するという基準により境界面を学習しており，また，一般に，カーネル関数を用いることにより非線形な境界面を学習することも可能である．複数モデルの出力を混合するタスクに対して SVM を適用する場合は，単語ラティス中の各単語の正誤を判別対象のクラス c とする．素性としては，以下の五種類を用いた(注1)．i) その単語を出力したモデルの情報，ii) その単語の品詞 (「茶筌」(注2) の最も粗い 9 品詞)，iii) その単語の音節数，iv) その単語を出力したモデルの情報とその単語の音響スコアをフレーム数で割ったスコアを結合したもの，v) その単語を出力したモデルの情報とその単語の言語スコアを結合したもの．SVM 学習・適用のツールとしては，Tiny-SVM (注3) を用いた．カーネル関数としては，多項式カーネルの一次および二次を評価したが，一次の方が性能がよかったので，4.2 節では，一次の多項式カーネルを用いた場合の評価結果を示す．また，`-c` オプション

(注1)：音響スコア・言語スコアについては，音響スコアをフレーム数で割らない場合，文全体の音響スコア・言語スコアを素性として用いた場合や素性として追加した場合など，様々な設定を評価したが，ここで述べる設定の性能を越えるような設定はなかった．

(注2)：http://chasen.aist-nara.ac.jp/

(注3)：http://cl.aist-nara.ac.jp/~taku-ku/software/TinySVM/

については，$1/((\text{訓練事例ベクトルの大きさ})^2$ の平均) を用いた．また，SVM の学習により得られた混合規則を適用する際には，評価事例と境界面との間の距離に下限を設け，単語ラティス上の競合する単語中で，これらの下限を満たす単語が存在すれば，その中で境界面からの距離が最も大きい高い単語を選択する．単語ラティス上の競合する単語中で，境界面からの距離の下限を満たす単語が存在しない場合には，その区間には高信頼度の単語は存在しないとして，単語を出力しない．

3.2 決定リスト学習

決定リスト [12] は，ある素性のもとでクラスを決定するという規則を優先度の高い順にリスト形式で並べたもので，適用時には優先度の高い規則から順に適用を試みていく．本論文では，各規則の優先度として，素性 f の条件のもとでの，クラス c の条件付き確率 $P(c \mid f)$ を用い，この条件付き確率順に決定リストを構成する．複数モデルの出力を混合するタスクに対して決定リスト学習を適用する場合は，単語ラティス中の各単語の正誤を判別対象のクラス c とする．決定リスト学習の素性としては，以下の二通りの設定を評価した(注4)．

i) 各単語を出力した二つのモデルの組を用いる

ii) 各単語を出力した二つのモデルの組，その単語の品詞 (「茶筌」の最も粗い 9 品詞)，その単語の音節数を用いて，モデル組，モデル組・品詞の結合，モデル組・音節数の結合，モデル組・品詞・音節数の結合の四種類の素性を作成し，この全素性を用いる．

決定リストの適用の際には，決定リストの各規則に対して，素性 f の条件のもとでのクラス c の頻度 $freq(f,c)$ の下限，および，条件付き確率 $P(c \mid f)$ の下限を設け，単語ラティス上の競合する単語中で，これらの下限を満たす単語が存在すれば，その中で最も優先度 (条件付き確率 $P(c \mid f)$) の高い単語を選択する．単語ラティス上の競合する単語中で，頻度および条件付き確率の下限を満たす単語が存在しない場合には，その区間には高信頼度の単語は存在しないとして，単語を出力しない．

4. 実験および評価

4.1 評価手順

本節では，前節で述べた機械学習手法 (SVM および決定リスト学習) により複数モデルの認識結果を混合する実験を行った評価結果について述べ，その単語認識率を，(重み付き) 多数決による混合の単語認識率と比較する．ただし，重み付き多数決においては，各モデルの認識率を重みとして用い(注5)，重みなし多数決においては，モデル数が同数の場合にその区間では単語を出力しない，という方法をとっている．混合の対象となる複数モデルとしては，2.1 節で述べた全 26 種類のモデルのうち，認識率の高い順に n 個 $(3 \leqq n \leqq 26)$ 選択して混合を行った．なお，単語認識率の評価においては，以下の尺度を用いる(注6)．

$$\text{再現率} = \frac{\text{正解単語数}}{\text{正解文の単語数}} \qquad \text{適合率} = \frac{\text{正解単語数}}{\text{認識結果の単語数}}$$

$$\text{F 値}_{\beta=1} = \frac{2}{\frac{1}{\text{再現率}} + \frac{1}{\text{適合率}}}$$

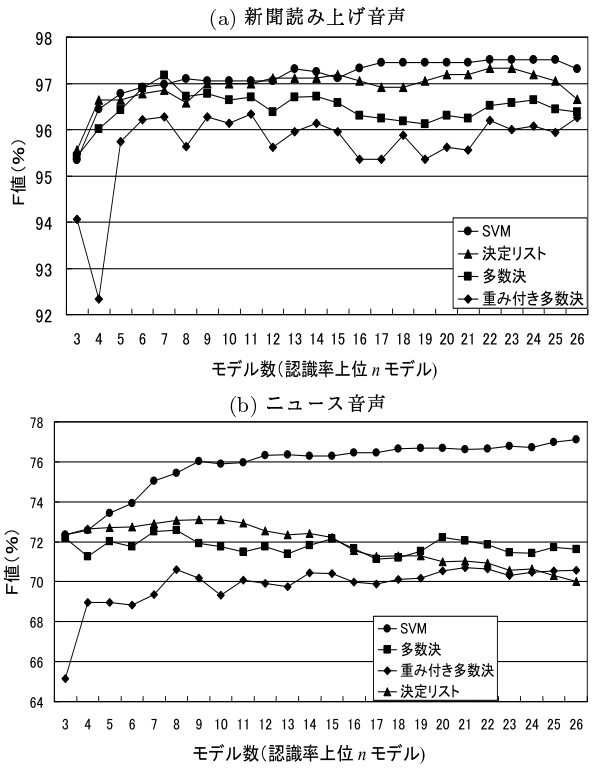

図 1: 認識率上位 n $(3 \leqq n \leqq 26)$ モデルの出力の混合における混合手法の性能比較

4.2 評価結果

4.2.1 混合手法の性能比較

まず，複数モデルの出力を混合する手法として，SVM によって学習した混合規則を用いる場合，決定リスト学習によって学習した混合規則を用いる場合，および，(重み付き) 多数決による場合の単語認識率を比較した．ただし，SVM においては，評価事例と境界面との間の距離の

(注4)：決定リスト学習における素性の設定においては，複数モデルの出力の混合において，ある単語が出力として選択されるためには，少なくとも二つ以上のモデルによって出力される必要がある，という制約を課している．前節で説明した SVM の素性の場合には，このような制約を課していないが，このような制約を課した素性の設定のもとで，SVM により複数モデルの出力の混合を行う評価実験も行った．この評価実験の結果においても，4.2 節の評価結果と同様に，SVM によるモデル混合の性能が決定リスト学習によるモデル混合の性能を上回った．

(注5)：文ごとのスコアを利用した重み付き多数決による混合の性能の評価も行った．具体的には，各単語について，その単語が含まれる文に各モデルが付与するスコアを求め，この文スコアからその文全体の単語認識率の推定値を算出し，この値を，その文中に含まれる各単語の重みとした．評価実験結果においては，この重み付き多数決法は，単純に各モデルの単語認識率を重みとする重み付き多数決法よりも高い性能を示したが，重みなし多数決法の性能は上回らなかった．ただし，データ整理の都合上，以下で示す重み付き多数決法による混合の評価結果は，全て，単純に各モデルの単語認識率を重みとした場合のものである．

(注6)：本論文では，ここで導入した適合率を信頼度とする評価尺度 [9] との整合性をとるために，再現率・適合率・F 値 $_{\beta=1}$ を用いて，混合結果の単語認識率を評価する．なお，ここでの再現率は単語正解率と同じ評価式となっている．

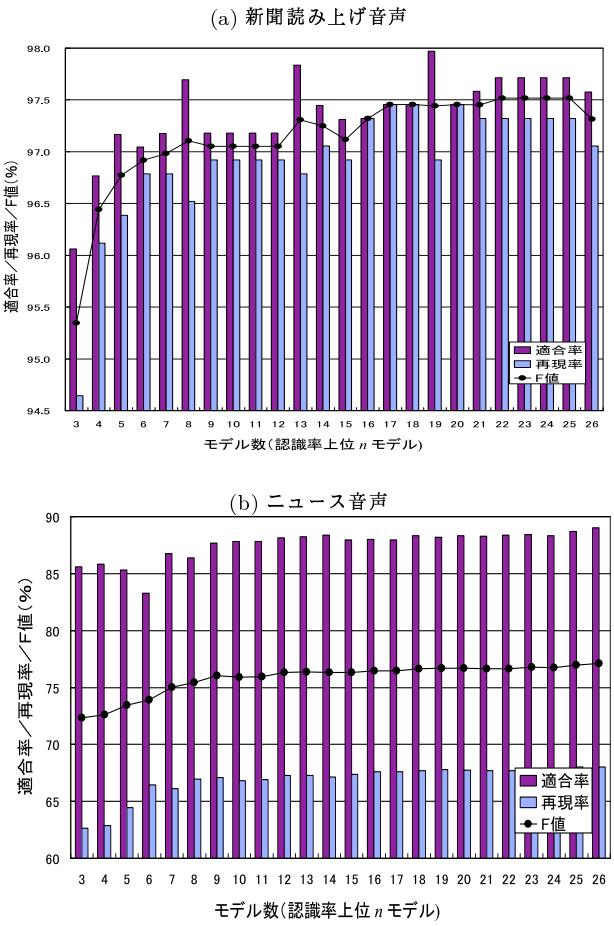

図 2: SVM を用いた認識率上位 n $(3 \leqq n \leqq 26)$ モデルの出力の混合における適合率・再現率・F 値の変化

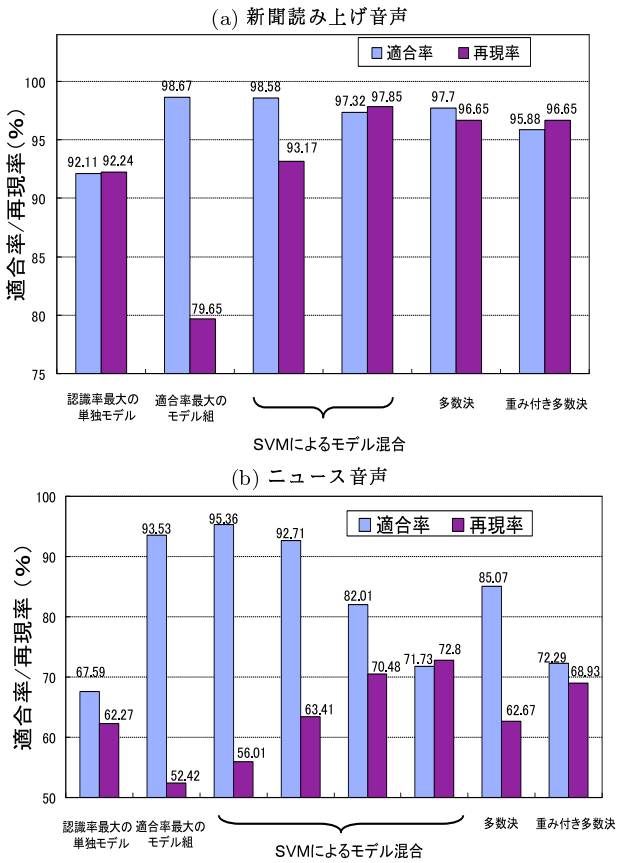

図 3: SVM による混合・(重み付き) 多数決による混合・単独モデル・二モデルの出力の共通部分の性能比較

下限の値に応じて，単語認識率の再現率および適合率のトレードオフが生じ，また，決定リスト学習においても，素性 f の条件のもとでのクラス c の頻度 $freq(f, c)$ の下限，および，条件付き確率 $P(c \mid f)$ の下限の値に応じて，単語認識率の再現率および適合率のトレードオフが生じる．そこで，これらの下限値については，単語認識率の F 値が最大となる値を用いることにする．この条件のもとで，図 1 に，混合対象となるモデルの数 n を変化させた場合の，各混合手法の F 値の変化を示す．ただし，各手法間で，素性等の利用する情報を合わせるために，SVM の素性として 3.1 節 iv) の音響スコア，および，v) の言語スコアを用いない場合の性能を，また，決定リスト学習の素性としては，3.2 節 ii) の設定を用いた場合の性能を，それぞれ示す．この結果から分かるように，新聞読み上げ音声の場合は，性能の高い順に，SVM，決定リスト学習，多数決，重み付き多数決，となった．ニュース音声の場合は，決定リスト学習の性能が悪くなっているが，次節で述べるように，素性としてモデル組のみを用いた場合は，SVM と多数決の間の性能を達成している．以上の結果から，SVM によって学習した混合規則を用いて複数モデルの出力の混合を行う手法が最も高い性能を達成していることが分かる．特に，決定リスト学習や (重み付き) 多数決法では，$n = 26$ よりも少ないモデル数の段階で F 値が最大値に達し，その後，モデル数が増えるにしたがって，混合結果の単語認識率 (F 値) が低下する傾向にある．したがって，決定リスト学習や (重み付き) 多数決による混合においては，性能の低いモデルが混入することにより，混合結果の単語認識率が低下する傾向があることが分かる．一方，SVM の場合でも，$n = 26$ よりも少ないモデル数の段階で F 値が最大値に達し，その後，モデル数が増えても，大幅な単語認識率の低下は起こらない．このことから，SVM を用いた混合では，性能の低いモデルが混入しても，学習の段階でこれに対処することが実現できていることが分かる．さらに，図 2 には，SVM を用いた混合について，混合対象となるモデルの数 n を変化させて，適合率・再現率・F 値の変化をプロットした結果を示す．この場合も，$n = 26$ よりも少ないモデル数の段階で適合率・再現率が十分高くなり，その後，モデル数が増えても，適合率・再現率の大幅な低下は起こらない．

次に，図 3 において，SVM による複数モデルの混合の単語認識率の再現率・適合率を，単語認識率 (再現率) 最大の単独モデル，(重み付き) 多数決による混合結果，などによる単語認識率 (再現率・適合率) と比較する．また，二つのモデルの出力の共通部分について，適合率が最大となるモデル組を求め [9]，そのモデル組の出力の共通部分の単語認識率 (再現率・適合率) も示す．ただし，SVM および (重み付き) 多数決については，混合対象となるモデルの数 n を変化させ，最も性能が良かった結果を示す．

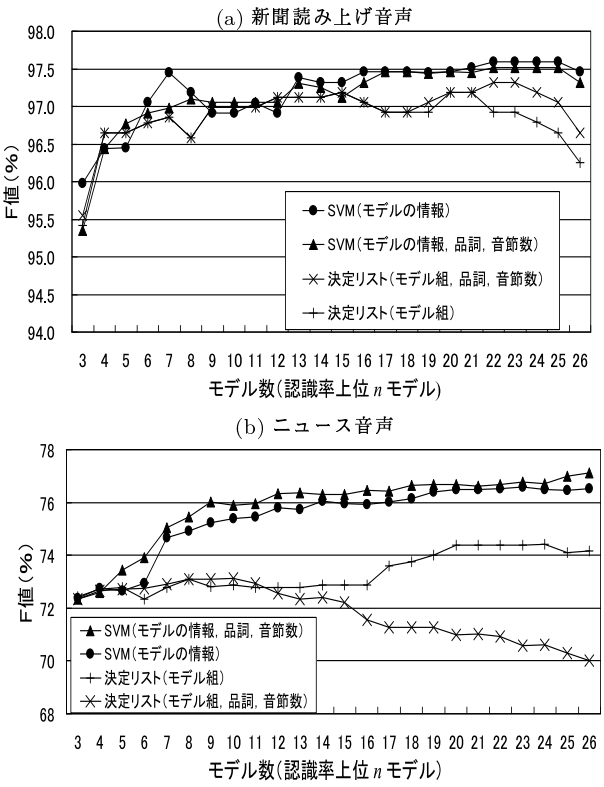

図 4: 認識率上位 n $(3 \leqq n \leqq 26)$ モデルの出力の混合における素性情報の性能比較

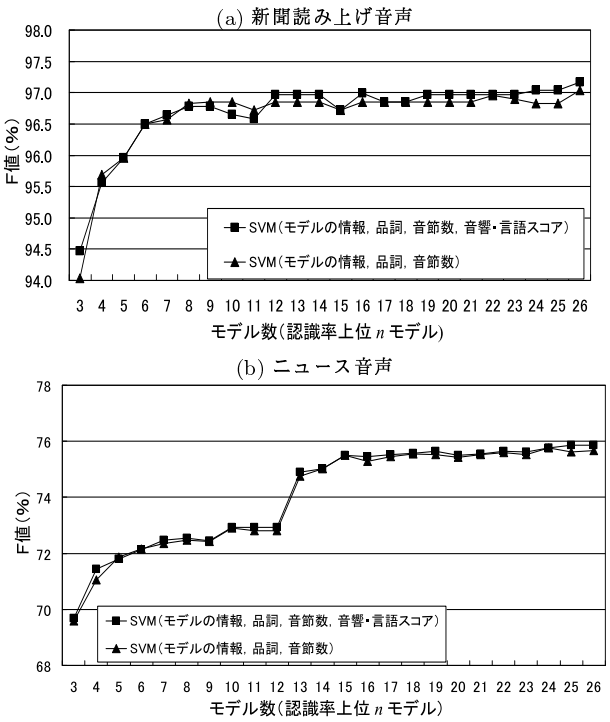

図 5: SVM を用いた認識率上位 n $(3 \leqq n \leqq 26)$ モデルの出力の混合における音響・言語スコアの評価

さらに，SVM については，評価事例と境界面との間の距離の下限の値を変化させ，比較対象との間で適合率と再現率の両方が比較できるような結果を何通りか示す．

この結果から分かるように，認識率最大の単独モデルとの比較では，適合率・再現率ともに大幅に改善できていることが分かる．また，適合率最大のモデル組と比較すると，新聞読み上げ音声では最大の適合率からはやや劣るものの，ニュース音声では最大の適合率を上回っている．また，再現率はいずれの場合も改善できている．特に，新聞読み上げ音声においては，適合率の劣化はごくわずかであるのに対して，再現率を大幅に改善している．また，(重み付き) 多数決との比較においても，適合率・再現率の両方においてほぼこれを上回る性能が達成できていることが分かる．特に，本論文の SVM の学習・適用による混合においては，評価事例と境界面との間の距離の下限の値を最適に調整することにより，適合率 (すなわち信頼度) あるいは再現率 (すなわちカバー率) のいずれかを優先した混合結果を出力することが容易に実現可能である．また，新聞読み上げ音声においては，認識率最大の単独モデルの再現率 92.24%が，SVM によるモデル混合により 97.85%に改善したので，その単語誤り改善率は，約 72%である．一方，ニュース音声においては，認識率最大の単独モデルの再現率 62.27%が，SVM によるモデル混合により 72.80%に改善したので，その単語誤り改善率は，約 39%である．同様に，多数決法との比較では，新聞読み上げ音声では約 36%の単語誤り改善率，および，ニュース音声では約 14%の単語誤り改善率を達成している．本論文の手法は，そのドメインの音声データおよび書き起こし正解文を数百文程度用意することができれば，どのようなドメインにも適用可能である．したがって，この条件が満たされれば，ROVER 法のような (重み付き) 多数決を上回る混合性能が達成でき，しかも，信頼度やカバー率を柔軟に調節した混合が実現可能であると言える．

4.2.2 機械学習の素性の性能比較

SVM および決定リスト学習について，素性の性能の比較を行った．まず，SVM については，3.1 節の素性 i)(各単語を出力したモデルの情報) のみを用いた場合と，素性 i)∼iii)(モデルの情報，品詞，音節数) を用いた場合の性能を比較した．決定リスト学習については，3.2 節の素性の設定 i)(各単語を出力したモデルの組) および設定 ii)(モデル組，品詞，音節数のあらゆる組合せ) について性能を比較した．混合対象となるモデルの数 n を変化させて，素性の各設定における F 値の変化をプロットした結果を図 4 に示す．この結果から分かるように，SVM と決定リスト学習を比べると，SVM の方が素性の設定の違いの影響を受けにくく，利用可能な素性が限られている場合でも，効果的な学習を行えることが分かる．特に，新聞読み上げ音声とニュース音声を比較すると，より難しいタスクであるニュース音声の認識結果の混合において，決定リスト学習が素性の設定の影響を受ける度合が高くなっている．また，各単語の品詞や音節数の有効性については，SVM の場合は，ニュース音声の認識結果の混合において品詞・音節数の情報が有効であることが分かる．

次に，SVM を用いた混合について，素性として音響スコア・言語スコアを用いた場合と用いない場合の性能を比較する．具体的には，3.1 節の素性 i)∼iii)(モデルの情報，

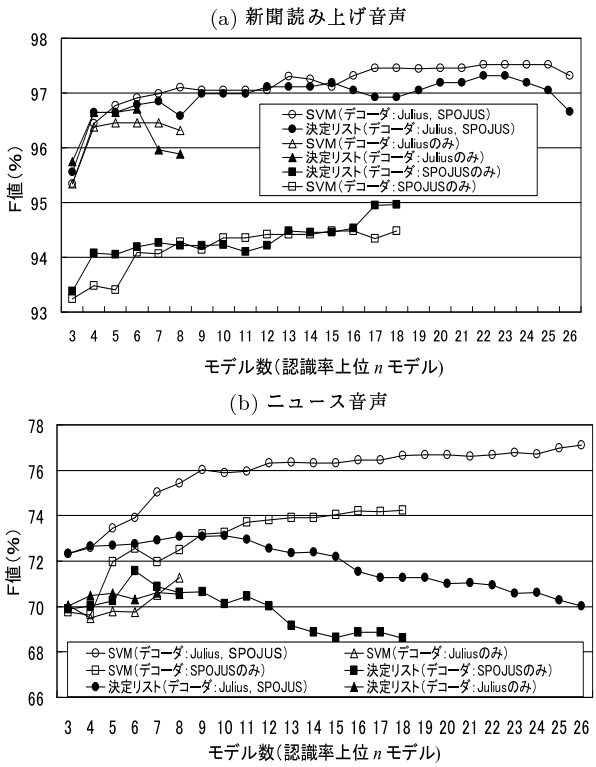

図 6: 認識率上位 n ($3 \leqq n \leqq 26$) モデルの出力の混合におけるデコーダの組合せの性能比較

品詞，音節数) を用いた場合と，i)∼v) の全素性 (モデルの情報，品詞，音節数，音響スコア，言語スコア) を用いた場合の性能を比較したものを図 5 に示す(注 7)．この結果から分かるように，音響スコア・言語スコアを用いて混合規則の学習を行うことにより，混合結果の単語認識率は向上するものの，その効果はわずかである．

4.2.3 デコーダの組合せの性能比較

最後に，複数の大語彙連続音声認識モデルの出力の混合において，異なるデコーダを用いた複数モデルの出力を混合する場合と，同一デコーダを用いた複数モデルの出力を混合する場合の性能の比較を行う．SVM および決定リスト学習の両方について，26 モデル全てに対して，認識率の高い順に n 個 ($3 \leqq n \leqq 26$) 選択して複数モデルの出力の混合を行った場合，デコーダが Julius の場合について，8 種類のモデルを認識率の高い順に n 個 ($3 \leqq n \leqq 8$) 選択して複数モデルの出力の混合を行った場合，および，デコーダが SPOJUS の場合について，18 種類のモデルを認識率の高い順に n 個 ($3 \leqq n \leqq 18$) 選択して複数モデルの出力の混合を行った場合の混合結果の単語認識率 (F 値) の変化を図 6 に示す．ただし，SVM の素性としては，3.1 節 iv) の i)∼iii) (モデルの情報，品詞，音節数) を用い，また，決定リスト学習の素性としては，3.2 節 ii) の設定を用いた．この結果から分かるように，SVM および決定リスト学習のいずれの場合も，異なるデコーダを用いた複数モデルの出力を混合する場合の性能の方が高くなっている．今回の実験では，デコーダが異なるモデルの方が，認識結果として異なる (正解) 単語を出力する傾向があり，複数モデルの出力の混合においても，個々のモデルができるだけ異なる (正解) 単語を出力した方が，混合結果の性能が高くなったものと考えられる．

(注 7)：図 4 の “SVM(モデル組，品詞，音節数)” の結果と図 5 の “SVM(モデル組，品詞，音節数)” の結果は，複数モデルの出力の混合の実験における設定は全く同じであるが，混合の対象となった単独モデルによる認識実験の設定が異なっていた．そのため，単独モデルの認識率の分布も異なったものとなっており，結果として，複数モデルの出力の混合の実験結果の単語認識率にも差が生じている．

5. おわりに

本論文では，機械学習を用いて複数の大語彙連続音声認識モデルの出力を混合するタスクに対して，SVM を適用した．評価実験の結果，認識率最大の単独モデル，および，決定リスト学習や (重み付き) 多数決を用いた混合の単語認識率を上回る性能が達成できた．

文　献

[1] 赤松裕隆, 花井建豪, 甲斐充彦, 峯松信明, 中川聖一. 新聞・ニュース文をタスクとした大語彙連続音声認識システムの評価. 情報処理学会第 57 回全国大会講演論文集, pp. 35–36, 1998.

[2] J. G. Fiscus. A post-processing system to yield reduced word error rates: Recognizer output voting error reduction (ROVER). In *Proc. ASRU*, pp. 347–354, 1997.

[3] K. Itou, et al. JNAS Japanese speech corpus for large vocabulary continuous speech recognition research. *The Journal of the Acoustical Society of Japan (E)*, Vol. 20, No. 3, pp. 190–206, 1999.

[4] 河原達也, ほか. 日本語ディクテーション基本ソフトウェア (99 年度版). 日本音響学会誌 (技術報告), Vol. 57, No. 3, pp. 210–214, 2001.

[5] T. Kemp and T. Schaaf. Estimating confidence using word lattices. In *Proc. 5th Eurospeech*, pp. 827–830, 1997.

[6] 中川聖一, 花井建豪, 山本一公, 峯松信明. HMM に基づく音声認識のための音節モデルと triphone モデルの比較. 電子情報通信学会論文誌, Vol. J83–D–II, No. 6, pp. 1412–1421, 2000.

[7] T. Utsuro, Y. Kodama, T. Watanabe, H. Nishizaki, and S. Nakagawa. Confidence of agreement among multiple LVCSR models and model combination by SVM. In *Proc. 28th ICASSP*, 2003. (to appear).

[8] 宇津呂武仁, 原田哲志, 渡邊友裕, 西崎博光, 中川聖一. 複数の大語彙連続音声認識モデルの出力の共通部分を用いた信頼度 — 信頼度を利用した複数モデルの出力の混合 —. 電子情報通信学会技術研究報告, SP2002-18∼23, pp. 25–30, 2002.

[9] 宇津呂武仁, 西崎博光, 原田哲志, 小玉康広, 中川聖一. 複数の大語彙連続音声認識モデルの出力の共通部分を用いた信頼度の性能分析. 電子情報通信学会技術研究報告, SP2001-125∼135, pp. 25–32, 2002.

[10] V. N. Vapnik. *The Nature of Statistical Learning Theory*. Springer-Verlag, 1995.

[11] F. Wessel, K. Macherey, and H. Ney. A comparison of word graph and N-best list based confidence measures. In *Proc. 6th Eurospeech*, pp. 315–318, 1999.

[12] D. Yarowsky. Decision lists for lexical ambiguity resolution: Application to accent restoration in Spanish and French. In *Proc. 32nd ACL*, pp. 88–95, 1994.