様式Ｃ－１９

科学研究費助成事業（科学研究費補助金）研究成果報告書

平成 24 年　5 月　31 日現在

機関番号：34310
研究種目：若手研究 B
研究期間：2009 ～ 2011
課題番号：21700184
研究課題名（和文）　認識単位が異なる認識器を併用した信頼度推定に関する研究

研究課題名（英文）　Automatic Speech Recognition with Confidence Measures Obtained by Multiple Recognizers of Various Recognition Units

研究代表者
西田 昌史（NISHIDA MASAFUMI）
同志社大学・理工学部・准教授
研究者番号：80361442

研究成果の概要（和文）：本研究では，カーナビゲーションシステムの目的地名を音声で設定する場面を想定し，目的地名が複合語であることに着目し認識単位の異なる認識器を 2 つ組み合わせて得られた信頼度を素性とした Support Vector Machine による認識結果の選択と棄却の枠組みを新たに実現した．多数の音声認識器を組み合わせる従来の手法に比べて，より少ない処理コストで高精度な音声認識を実現し，システムに登録されていない未知語も棄却することができた．

研究成果の概要（英文）：In this study, I proposed a novel speech recognition method that improves the word recognition accuracy by using confidence measures obtained from multiple recognizers of various recognition units. The method uses Support Vector Machine (SVM) to select a single recognition result from multiple recognition results. The method can also identify whether the correct recognition result is included in the multiple recognition results. Experimental results show that the proposed method gives higher word accuracy and classification accuracy in comparison with a single recognizer and ROVER.

交付決定額

（金額単位：円）

	直接経費	間接経費	合　計
２００９年度	400,000	120,000	520,000
２０１０年度	600,000	180,000	780,000
２０１１年度	500,000	150,000	650,000
年度			
年度			
総　計	1,500,000	450,000	1,950,000

研究分野：総合領域
科研費の分科・細目：情報学，知覚情報処理・知能ロボティクス
キーワード：音声認識，認識単位，複合語，信頼度，Support Vector Machine，音声対話

１．研究開始当初の背景

従来の音声認識では，音素単位で構成した音響モデルと単語単位で構成した N-gram に基づく言語モデルが一般的に用いられている．これらのモデルを組み合わせて，それぞれのモデルから算出された音響尤度と言語尤度を重み付けして統合することで認識を行っている．

しかしながら，従来のように音響情報と言語情報を尤度のレベルで統合しているため，音響的な特徴と言語的な特徴のどちらが主

に認識結果の誤りに影響を与えているのかを判断することは難しいと考えられる．つまり，音響的に類似した候補が存在しているからなのか，あるいは学習時の単語間の N-gram 確率，パープレキシティといった単語の出現割合が影響しているかである．これらの影響の違いによってモデルや認識器の構成，認識結果の改善手法が異なってくると考えられる．

さらに，日本語において複数の単語を組み合わせて構成される複合語（例えば，「同志社」と「大学」で「同志社大学」，「音声」と「認識」で「音声認識」など）が多く存在している．一般的に，N-gram による言語モデルを構成する際，単語単位にすることで多くの複合語を表現することが可能となるが，N-gram の場合認識を連続して誤ってしまう問題があり，複合語を構成するいずれかの単語を誤って認識してしまうと意味がまったく異なってしまう．また，単語が短いため音響的に類似する候補が増えてしまう可能性がある．

２．研究の目的

さきほど述べた問題点を解決するために，音素，単語，文節といったように認識単位の長さの異なるものを併用し，それらの認識結果などを統合することで信頼度を推定し，認識誤りを訂正する手法について検討を行う．音響情報のみを用いた認識と言語情報を加えたときの認識を独立して行うことで，それぞれの影響による信頼度を算出することができる．

また，従来の N-gram での単語単位だと短いため音響的な類似候補が多くなることに対して，従来の単語単位にさらに複合語単位でモデル化することで，音響的に類似する候補を減らした場合での認識についても検討を行う．

タスクとしては，カーナビゲーションシステムの目的地名設定を想定して音声認識の評価を行う．

３．研究の方法

（1）初年度は，地名を形態素解析し得られた形態素単位，地名全体（「同志社大学」など），地名を名称（「同志社」など）と属性（「大学」など）に分割した単位の 3 つのパターンで N-gram と FSA（文法）による音声認識器を構築する．

これらの認識器から得られる認識結果の信頼度を推定するための尺度として，単語事後確率，音響尤度，言語尤度，認識器間の認識結果の一致度と音節数差を用いる．

これらの信頼度尺度をもとに，認識単位の異なる認識器を 2 つずつペアにして，先ほどの信頼度尺度を特徴量として Support Vector Machine(SVM)により認識結果の選択を行う．

（2）二年度目は，昨年度では目的地名を名称と属性に手動で分割していたが，目的地名を形態素解析し得られた形態素間の連節確率に対して閾値を設けることで，形態素間を自動的に結合することで新たに部分単語単位 N-gram を生成する．

また，信頼度としては音響尤度，言語尤度，単語事後確率，N-best リスト中の一致度，認識結果間の音素数差を用いて，2 つの認識結果をもとにどちらの認識結果が正しいかどちらも正しくないかを SVM により判別する．

さらに，昨年度は認識対象の目的地名が 1 万個であったが，目的地名を 10 万個に増やしより大規模な評価実験を行う．

（3）三年度目は， 昨年度までで，正しい認識結果が得られなかった場合に認識結果を棄却する枠組みとして，認識単位が異なる認識器を 2 つ組み合わせて得られた信頼度を素性とした SVM による認識結果の選択と判別について検討を行った．しかし，評価データに含まれる地名はすべて辞書に含まれていたため，辞書に含まれていない地名を含めて未知語に対する評価を行う．

また，複数の認識器を組み合わせた従来手法としてよく用いられている ROVER 法との比較実験を音声認識ならびに未知語の棄却の評価に対して行う．今回の ROVER 法は，形態素単位 N-gram，部分単語単位 N-gram，単語単位 FSA の 3 つの認識器をすべて用いて，得られた認識結果に対して多数決処理を行う．その際，最も多い認識結果を選択し，すべての認識結果が異なる場合は認識結果を棄却することとする．

４．研究成果

(1) 初年度では，どの認識単位の認識器の組み合わせが最も認識結果の信頼度を推定するのに最適であるかを明らかにするために，認識単位の異なる認識器を 2 つずつペアにして，先ほどの信頼度尺度を特徴量として図 1 のように SVM による認識結果の成否を学習し判別する手法を提案した．

認識対象は 10,000 地名で，10 名の被験者が 100 種類の地名を 4 つの言い回しで発話した 4,000 データを実験に用い，交差検証法により学習データと認識データの組み合わせを変えて実験を行った．デコーダには Julius4.1.2 を用いた．

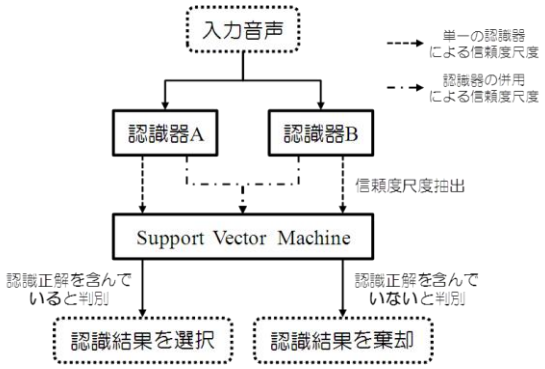

図 1 提案手法の処理の流れ

実験を行った結果，単一の認識器を用いた場合形態素単位 N-gram で 85.8%，名称・属性単位 N-gram で 87.4%，地名単位（単語単位）FSA で 88.1%の認識精度が得られた．それに対して，提案手法により形態素単位 N-gram と地名単位文法の組み合わせにおいて，89.6%の最も高い認識精度が得られた．したがって，通常の単体での認識器に比べて，認識単位の異なる認識器を組み合わせて得られた信頼度尺度により認識精度を改善することができ，提案手法の有効性を示すことができた．

（2）二年度目は，目的地名を形態素解析し得られた形態素間の連節確率に対して閾値を設けて，認識精度が最も高くなるように閾値を制御することで形態素間を自動的に結合することで部分単語単位 N-gram を生成することができた．図 2 に形態素間の連接確率に対する閾値を変えたときの音声認識精度を示す．

認識器としては，形態素単位 N-gram，部分単語単位 N-gram，単語単位 FSA の 3 つから 2 つを組み合わせて，得られた信頼度をもとに 2 つの認識結果のうちどちらが正しいかどちらも正しくないかを SVM により判別した．

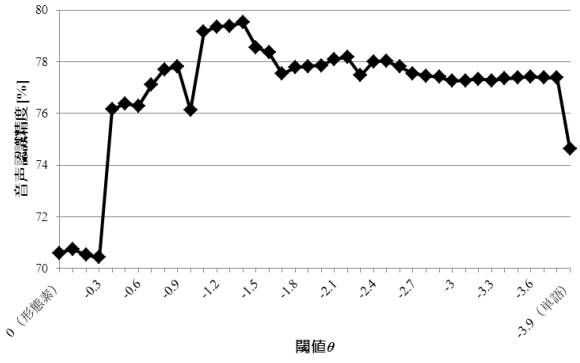

図 2 形態素間の連接確率に対する閾値を変化させたときの音声認識精度

昨年度に比べて目的地名を 10 万個に増やしより大規模な評価実験を行った．評価データは，10 名の被験者が 100 個の目的地名を 4 つの言い回しで発話した 4,000 データである．

その結果，単一の認識器では F 値が形態素単位 N-gram で 89.2，部分単語単位 N-gram で 91.9，単語単位 FSA で 93.2 という判別精度が得られた．それに対して，提案手法である 2 つの認識器を組み合わせた場合では F 値が形態素単位 N-gram と部分単語単位 N-gram で 93.1，部分単語単位 N-gram と単語単位 FSA で 96.9，単語単位 FSA と形態素単位 N-gram で 96.4 という結果が得られた．

以上の結果から，認識単位の異なる認識器を組み合わせた提案手法は，認識結果の正否判別において有効であることが明らかになった．

（3）三年度目は，評価データとして 4000 発話を用いそのうち 1200 発話に未知語を含めた．評価実験の結果，認識結果の分類精度は単一の認識器を用いた場合，形態素単位の N-gram で 81.4%、部分単語単位の N-gram で 79.7%，単語単位の FSA で 73.2%，3 つの認識器を組み合わせた ROVER 法で 69.8%という結果が得られた．

それに対して，提案手法では形態素単位の N-gram と部分単語単位の N-gram の組み合わせで 79.0%，部分単語単位の N-gram と単語単位の FSA の組み合わせで 83.1%，単語単位の FSA と形態素単位の N-gram の組み合わせで 81.5%という結果が得られた．

また，単一の認識器，ROVER 法，提案手法のそれぞれにおける音声認識精度を図 3 に示す．この結果から，提案手法は従来の単一の認識器や ROVER 法に比べて認識精度が改善されており，部分単語単位 N-gram と単語単位 FSA の組み合わせにおいて最も高い認識精度が得られた．

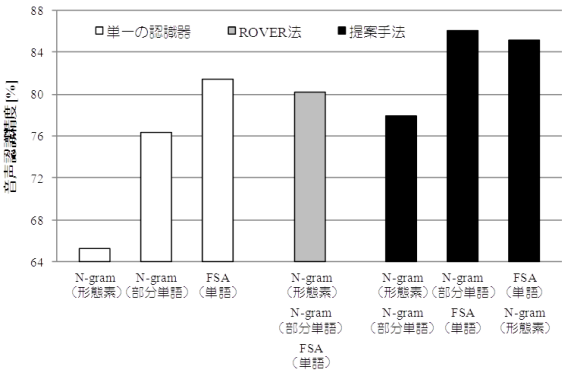

図 3 各手法による音声認識精度

したがって，提案手法は未知語に対しても高精度に棄却することができ，高い音声認識精度を実現できることが明らかになった．また，従来よく用いられてきた ROVER 法では，一般的に多くの音声認識器を用いているため処理コストがかかっていたが，提案手法は

2 つの音声認識器を用いるだけで済むため，処理コストを大幅に抑えることも可能である．

今回は，認識単位の異なる認識器を組み合わせて得られた信頼度により音声認識精度の改善ならびに認識結果の棄却を実現することができたが，どのような評価データに対してどの認識器の組み合わせがよいかなど詳細な分析を今後行っていきたいと考えている．

また，今回は 2 つの認識器の認識結果の第一位候補に正解がなければ棄却するという枠組みであったため，第二位候補以降に正解が含まれていても認識精度を改善することができなかった．そこで，今後は認識結果の第二位候補以降も考慮した枠組みについても検討を行っていきたいと考えている．

５．主な発表論文等

（研究代表者、研究分担者及び連携研究者には下線）

〔雑誌論文〕（計 1 件）

① N. Suzuki, T. Tanaka, M. Nishida, S. Yamamoto, Selection and Rejection of Recognition Results Obtained with Multiple Recognizers of Various Recognition Units Using Support Vector Machine, Journal of Signal Processing, 査読有, Vol. 16, No. 4, pp. -, 2012.（掲載決定）

〔学会発表〕（計 6 件）

① Nobuhisa Suzuki, Automatic Speech Recognition with Confidence Measures Obtained by Multiple Recognizers of Various Recognition Units, International Workshop on Nonlinear Circuits, Communications, and Signal Processing (NCSP), 2012 年 3 月 4 日, Waikiki Beach Marriott Resort & Spa (Honolulu).

② 西田 昌史, 異なる認識単位の認識器を併用した SVM による認識結果の選択と判別, 日本音響学会 2011 年春季研究発表会, 2011 年 3 月 10 日, 早稲田大学（東京）.

③ 田中 智之, 異なる認識単位の認識器から得られた信頼度を素性に用いた音声認識, 第 12 回音声言語シンポジウム, 2010 年 12 月 21 日, 国立オリンピック記念青少年総合センター（東京）.

④ 田中 智之, 認識単位の異なる認識器から得られた信頼度に基づく音声認識, 日本音響学会 2010 年秋季研究発表会, 2010 年 9 月 15 日, 関西大学（大阪）.

⑤ Masafumi Nishida, Automatic Speech Recognition Based on Multiple Level Units in Spoken Dialogue System for In-vehicle Appliances, 13th International Conference on Text, Speech and Dialogue(TSD), 2010 年 9 月 7 日, Hotel Continental in Brno (Czech Republic).

⑥ 田中 智之, 複合語を対象とした異なる認識単位による音声認識の検討, 日本音響学会 2010 年春季研究発表会, 2010 年 3 月 8 日, 電気通信大学（東京）.

６．研究組織

(1)研究代表者

西田 昌史（NISHIDA MASAFUMI）

同志社大学・理工学部・准教授

研究者番号：80361442