

頑健な意味理解のための音声認識信頼度と対話履歴を利用した 発話意図推定手法

水野 智士[†], 高木 浩吉[†], 小暮 悟[†], 甲斐 充彦^{††}, 伊藤 敏彦[‡], 小西 達裕[†], 伊東 幸宏[†]

[†] 静岡大学情報学部 ^{††} 静岡大学工学部

[‡] 北海道大学情報科学研究科

E-mail: [†] {cs9082,cs0054,kogure,konishi,itoh}@cs.inf.shizuoka.ac.jp, ^{††} kai@sys.eng.shizuoka.ac.jp,
[‡] t-itoh@media.eng.hokudai.ac.jp

あらまし 近年の音声認識、言語理解技術、及びコンピュータ性能の向上によって、音声を用いるインタフェースやタスク指向型の対話システムが利用されるようになってきた。そんな中で、より一般的にシステムが利用されるようになるには、より頑健な言語理解が必要となる。本稿では、より頑健な意味理解を実現するために、音声認識信頼度と対話履歴を利用して、ユーザ発話意図の推定を行う手法について記述する。本研究では、言語理解の頑健さを向上させるために、対話履歴において、県名や市町村名など、どのカテゴリについての発話が行われたのかを識別する。その識別結果と、認識結果の n-best を利用して言語理解結果を生成する。これを実現する場合、カテゴリ識別の精度がそのまま言語理解精度に影響する。そこで、ユーザの発話意図を推定することで、カテゴリ識別精度の向上を図り言語理解精度向上を目指した。評価実験を行い、音声認識の 1-best をそのまま利用する言語理解手法よりも提案手法のほうが、言語理解精度が高くなることを示した。

Speech intent presumption method using confidence score of speech recognition and dialogue history for robust meaning understanding

Satoshi Mizuno[†], Hiroyoshi Takagi[†], Satoru Kogure[†], Atsuhiko Kai^{††},

Toshihiko Itoh[‡], Tatsuhiro Konishi[†], Yukihiro Itoh[†]

[†] Faculty of Informatics, Shizuoka University ^{††} Faculty of Engineering, Shizuoka University

[‡] Graduate School of Informatics, Hokkaido University

E-mail: [†] {cs9082,cs0054,kogure,konishi,itoh}@cs.inf.shizuoka.ac.jp, ^{††} kai@sys.eng.shizuoka.ac.jp,
[‡] t-itoh@media.eng.hokudai.ac.jp

Abstract The spoken dialogue interface and the task oriented dialogue system has come to be used by improving the speech recognition, the language understanding technologies, and the computer performance. We need a more robust language understanding for the system to come to be used more generally. Our paper deals with speech intent presumption method using the confidence score of speech recognition and dialogue history for robust meaning understanding. This language understanding results are generated by using the speech recognition results (n-best) and the identification results. Thus, the accuracy of the category identification influences the language understanding accuracy. Then, we used the presumption of user's speech intention in order to improve the language understanding accuracy. As the result of evaluation experiment, we show that the language understanding performance used our proposed method is higher than the language understanding method which simply gives priority to the first hypothesis of an n-best.

1. はじめに

近年の音声認識技術や言語処理技術及びコンピュータ性能の向上などにより、音声インタフェースや目的指向型音声対話システムが注目され始めている。音声インタフェースの利

用は、様々な環境・対象に広がっており、特にカーナビゲーションシステムにおいて音声操作が可能になってきている。運転という主となる行為と平行する操作としては、「目」や「手」を使わなくてはならないリモコン操作よりも、音声入

力操作の方が安全である。しかしながら、自然発話であることや走行ノイズ等の影響により、現在の音声認識技術では誤認識を回避することは困難である。ユーザの発話した文を誤認識すると、ユーザの発話を正しく理解することが出来ず、対話の流れが不自然になる可能性がある。その結果、正常に対話が進行した場合に比べてユーザが不快感を覚えてしまう。このような現実をふまえ、音声認識の誤認識に対処するための様々な研究が行われている[1][2]。音声認識結果の信頼度を利用する手法[3][4]についても研究されているが、信頼度のみを利用しただけでは、誤認識に対して十分に対処することは困難である。

そこで我々は、誤認識した場合において、信頼度に加えて、対話履歴などの情報を用いることによって、システム自身が誤認識を正しく判断し修正できる頑健な音声対話システムの構築を目指している[5][6]。

本稿では、まず、音声認識の信頼度と対話履歴を利用して、ユーザの発話意図の推定を行う。それにより言語理解率を向上させる。選択したタスクは、目的地設定タスク、車載機器操作タスクの2つのタスクである。言語理解率による評価には、音声認識結果(1-best)を最優先に用いたシステム、クラススコアを用いたカテゴリ理解を行う言語理解手法[5]を用いたシステム、及び本稿で紹介する手法を用いたシステムの3つのシステムを用いる。これら3つのシステムの評価実験の結果を比較し、本提案手法の有効性を示す。

2. 発話タイプとタスク

提案手法は様々なタスクに対応できるが、本研究では目的地設定タスクと車内機器操作タスクの2つのタスクを選択した。発話タイプとそれぞれのタスクについて説明する。

2.1. 発話タイプ

ユーザの発話は以下に示すとおり、詳細化、訂正、回答、再入力の4つの発話タイプに分類できる。後述の目的地設定タスクの対話例とそのときのユーザ発話の発話タイプを図1に示す。

- **詳細化** 応答された内容に対して情報を追加する発話
- **訂正** 応答された内容に対して訂正を行う発話
- **回答** 質問を含む応答に対して回答する発話
- **再入力** 再入力要求の応答に対する発話

2.2. 目的地設定タスク

目的地設定タスクとは、カーナビゲーションシステムの目的地(ランドマーク[以下、LM])設定を行うタスクである。LMは、インター名、駅名、市区町村名を指す。これらを木構造になるよう、図2のようにPR、HR、LMに分類する。この分類のことをカテゴリと呼ぶ。また、カテゴリの各要素をクラスと呼ぶ。

U1: 静岡県の
 S1: 静岡県
 U2: 浜松西IC (詳細化発話)
 S2: 静岡県の何ICですか
 U3: 浜松西ICです (回答発話)
 S3: 静岡県の浜松ICですか
 U4: いいえ、浜松西です (訂正発話)
 S4: もう一度発話してください
 U5: 浜松西ICです (再入力発話)
 S5: 静岡県の浜松西ICですか
 U6: はい
 S6: 目的地に設定しました

図1: 目的地設定タスクの対話例

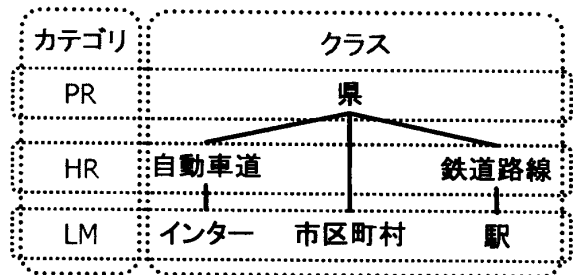


図2: 目的地設定タスクの発話内容の分類

ユーザ発話は、一度で全てのカテゴリを入力、または複数回に分けて入力できる。さらに、「はい」や「いいえ」などの肯定語や否定語も発話することが可能である。

2.3. 車内機器操作タスク

車内機器操作タスクとは、車内のエアコンやカーステレオの操作を行うタスクである。車内機器操作タスクでの発話可能な単語を目的地操作タスク同様に木構造になるように分類したものを図3に示す。

ユーザは「対象を操作」、「属性は値」という文体で発話可能である。また、目的地設定タスクと同様に肯定語や否定語を発話することが可能である。

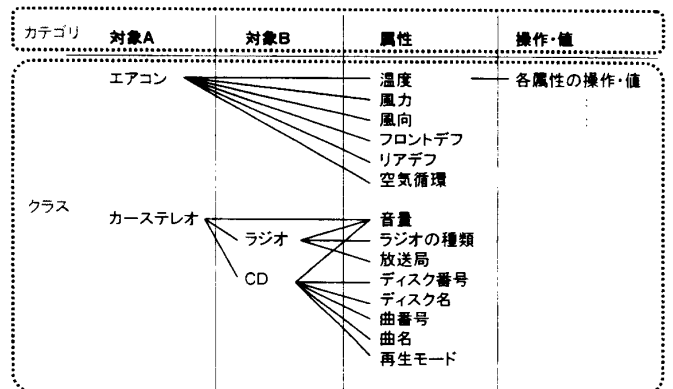


図3: 車内機器操作タスクの発話内容の分類

3. 言語理解

図4にシステムの構造図を示す。システムは、音声認識部、信頼度生成部、言語理解部、応答生成部、GUI表示部、音声合成部から成る。各モジュール間のデータの受け渡しはTCP/IPを利用している。ただし、応答生成部と音声合成部の間のみクリップボードによる受け渡しを行っている。

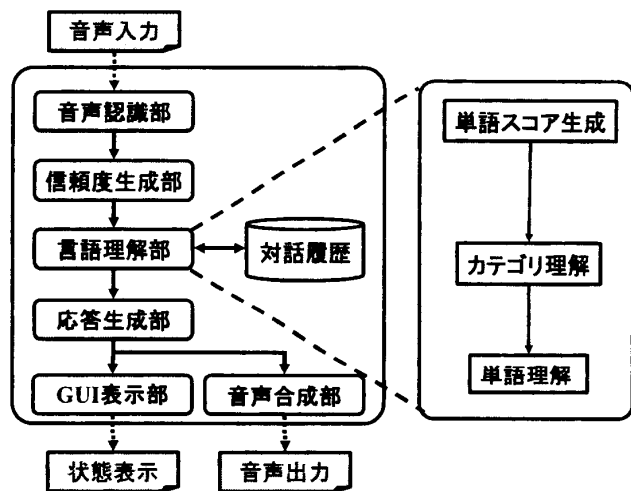


図4: システムの構成

3.1. 音声認識部

音声認識部は、入力音声を音声認識し、CFGによる文法制約の元で音響尤度によって順序付けされた複数候補からなる認識結果(n-best)を生成する。生成されたn-bestは信頼度生成部へ入力される。本システムでは、音声認識部としてSPOJUS[7]を用いる。SPOJUSは豊橋技術科学大学の中川研究室で開発された、UNIXまたはWindows上で動作する、日本語連続音声認識システムである。

3.2. 信頼度生成部

信頼度生成部では、音声認識部で生成されたn-bestをもとに、単語/クラスのそれぞれについて信頼度を生成する。この音声認識結果の信頼度は、音響的な尤度とn-best中の存在率から事後確率に基づく尺度として計算される[3]。信頼度の付与されたn-bestは、言語理解部に入力される。

3.3. 言語理解部

言語理解部では、信頼度生成部で生成された信頼度付きn-bestと対話履歴を用いて単語スコアを生成する。次に、単語スコアからカテゴリ理解を行い、最後に単語理解が行われることで言語理解内容が生成される。

単語スコアとは、ある単語がこれまでの文脈の中で、どれくらい発話されているかの可能性を示す値である。生成された単語スコアは、履歴に基づく総合認識結果として対話履歴に残される。

3.3.1. 単語スコア生成

単語を(1)対話履歴中の単語と(2)最新の認識結果中の単語の2つに分類し、それぞれ異なる戦略を用いてスコアを生成する。(2)の単語は、最新認識結果の複数の候補(n-best)に含まれる全単語が対象となる。単語スコア生成は、言語理解部が最新の認識結果を獲得するたびに、(1)(2)の順に単語スコア生成が行われ、(2)の単語スコアは対話履歴中の単語スコアへと更新される。この時(1)(2)の両方でスコアを生成された単語については、高いスコアの方を対話履歴中の単語スコアに更新する[5]。

3.3.2. カテゴリ理解

カテゴリ理解とは、対話履歴中の単語スコアを用いてユーザがこれまでに発話した内容をカテゴリレベルで理解を行うことである。例えば目的地設定タスクで『静岡県浜松IC』と発話した場合に正しくカテゴリ理解を行うと、カテゴリ理解結果は「PRとLM」となる。

カテゴリ理解の手順は、まず、(A)対話履歴中の単語スコアを用いて「PR」や「PRとHR」などのカテゴリの全組み合わせに、その組み合わせのカテゴリがこれまでの文脈の中でどれくらい発話されているかの可能性を示す値としてのスコア(以下、カテゴリスコア)をそれぞれ求める。例えば目的地設定タスクにおいては、カテゴリ数は3なので、その全組み合わせは、 ${}_3C_1 + {}_3C_2 + {}_3C_3 = 3 + 3 + 1 = 7$ 通りとなり、カテゴリの組み合わせそれぞれにカテゴリスコアを求める。カテゴリスコアの求め方は以下で説明する。

次に、(B)対話履歴の単語スコアや前回の応答内容を参考に、最新の発話が前回の応答に対してどのような発話タイプ(詳細化や回答など)であるのかを各カテゴリで判断する。ただしこの発話タイプの判別で、判別基準に満たない場合は、判別不能という状態とする。(C)発話タイプが詳細化や回答などに判別されたカテゴリを含むカテゴリの組み合わせには、(A)で計算されたカテゴリスコアの判別結果に沿った増減を行う。このカテゴリスコアの増減によってカテゴリスコアの精度を上げる。発話タイプの判別方法や増減方法などは以下で説明する。

最後に、(D)カテゴリの全組み合わせの中から最も高いカテゴリスコアを持つカテゴリの組み合わせをカテゴリ理解結果とする。

このカテゴリ理解の(A)から(D)の流れを、目的地設定タスクを例として詳しく説明する。

(A) 単語スコアを用いて「PR」や「PRとHR」などの全組み合わせのカテゴリのスコアを求める。

ここでは、カテゴリの全組合せの、発話された可能性を示す値であるカテゴリスコアを対話履歴中の単語スコアから計算で導く。カテゴリスコアを求めるカテ

ゴリの組み合わせにおいて、「PR」や「LM」などを単独カテゴリパターンと「PRとLM」や「PRとHRとLM」などを複数カテゴリパターンと呼ぶ。

まず単独カテゴリパターンでは、各カテゴリ内の最も高い単語スコアをそのカテゴリのスコアとする(図5)。

単語	スコア	カテゴリ	合計値
静岡	0.75	PR	0.80
神奈川	0.80	HR	1.57
<hr/>			
東名自動車道	1.57	LM	0.88
浜松西IC	0.88	PR+HR	
浜松IC	0.32	PR+LM	
浜松市	0.25	HR+LM	
		PR+HR+LM	

図5:カテゴリスコアの決定手順1

次に複数カテゴリパターンでは、単に単独カテゴリパターンのカテゴリスコアを加算するのではなく、地理制約を考慮して計算する。例えば図5の場合では、複数カテゴリパターン「PRとLM」のカテゴリスコアを計算するとき、地理を無視した「神奈川」と「浜松西IC」の両単語スコアの合計値を用いるのではなく、「静岡」と「浜松西IC」の両単語スコアの合計値を用いる。このとき、直接比較できるように、それぞれのカテゴリパターンに含まれるカテゴリ数に応じて、合計スコアをパラメータを用いて正規化する。なおパラメータは2カテゴリパターンでは1.6を、3カテゴリパターンでは2.2を用いた。これらのパラメータや(B)で用いる閾値はヒューリスティックによって求めた(図6)。

単語	スコア	カテゴリ	合計値
静岡	0.75	PR	0.80
神奈川	0.80	HR	1.57
<hr/>			
東名自動車道	1.57	LM	0.88
浜松西IC	0.88	PR+HR	1.48
浜松IC	0.32	PR+LM	1.01
浜松市	0.25	HR+LM	1.53
		PR+HR+LM	1.45

図6:カテゴリスコアの決定手順2

(B) 対話履歴の単語スコアや前回の応答内容を参考に、最新の発話が前回の応答に対してどのような発話タイプ(詳細化や回答など)であるのかを各カテゴリで判断する。

ここでは、対話履歴の単語スコアや前回答の内容を参考に、最新発話がどのような発話タイプであるかを各カテゴリで調べ、その結果に沿ってカテゴリスコアの微調整を行い、精度を上げる。発話タイプの判別は図7のフローチャートに沿って行われる。この判別方法によって各カテゴリ全てがどの発話タイプにも属しないと判断された場合は、カテゴリスコアの微調整を行わず、(D)の理解内容のカテゴリの決定を行う。判別結果にどの発話タイプにも属さない結果を作成した

理由は、4つの発話タイプに全てのカテゴリを判別すると、発話タイプの判別が誤ったときにカテゴリスコアの精度に悪影響を及ぼす恐れがあるからである。

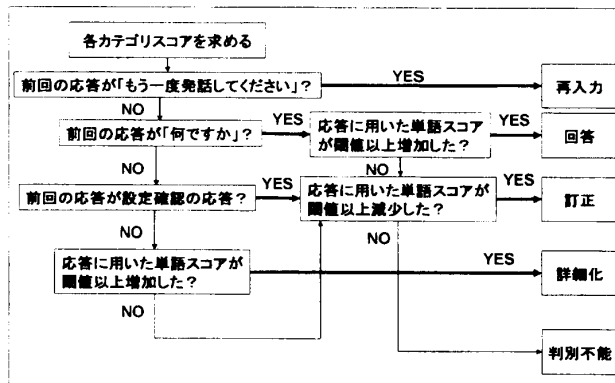


図7:発話タイプ判別のフローチャート

(C) 発話タイプが詳細化や回答などに判別されたカテゴリを含むカテゴリの組み合わせには、計算されたカテゴリスコアの判別結果に沿った増減を行う。

ここでは、発話タイプの判別結果に沿った処理によって、カテゴリスコアの微調整を行う。それぞれの発話タイプの処理について以下で説明する。

☆詳細化

カテゴリが詳細化と判断された場合には、詳細化されたカテゴリと他のカテゴリを含む複数カテゴリパターンのカテゴリスコアに『詳細化された可能性を示す値』を加える。これによって、詳細化されたカテゴリを忘れず、かつ詳細化したカテゴリを追加することが出来る。この『詳細化された可能性を示す値』は、応答に使用した単語スコアの増加値(「応答に使用した単語の対話履歴の単語スコア」-「応答に使用した単語の前回までの対話履歴の単語スコア」)で求める。

☆回答

カテゴリが回答と判断された場合には、詳細化と同様の処理を行う。回答と詳細化は、システムが応答で詳細化するカテゴリ(クラス)を指定する(回答)か、しない(詳細化)かだけが違うからである。

☆訂正

カテゴリが訂正と判断された場合には、まず、訂正と判断されたカテゴリ自体が間違っているのか、カテゴリは正しいが単語内容が間違っているのかを判断し、カテゴリが間違っているときにはそのカテゴリを含むカテゴリパターンのカテゴリスコアを減少させ、カテゴリが正しいときにはそのカテゴリを含むカテゴリパターンのカテゴリスコアを増加させる必要がある。これを実現するために『訂正された可能性を示す値』を加えることにする。この『訂正された可能性を示す値』は、以下のように求める。

●訂正と判断されたカテゴリ内で前回応答と同じ単語の単語スコアが一番高いとき(カテゴリが間違っている時)

応答単語の前回までの対話履歴の単語スコアと最新の対話履歴の単語スコアの差分

●訂正と判断されたカテゴリ内で前回応答と異なる単語の単語スコアが一番高いとき(カテゴリは正しい時)

訂正と判断されたカテゴリ内の最もスコアの高い対話履歴の単語スコアと前回応答の単語の対話履歴の単語スコアの差分

☆再入力

カテゴリが再入力と判断された場合には、そのカテゴリを含むカテゴリパターンのカテゴリスコアに『再度発話された可能性を示す値』を加える。この『再度発話された可能性を示す値』は、前発話での各カテゴリの最高単語スコアが今回の発話によって増減した値を用いる。この値は、前発話の最高スコアを持つ単語が発話された単語なら今回の発話によってスコアが上昇し、逆に前発話の最高スコアを持つ単語が発話されていない単語なら今回の発話によってスコアが減少、または少ししか上昇しないはずである。これによって再度発話された内容のカテゴリを理解しやすくする。

(D) 各カテゴリスコアの中から最も高いスコアを持つカテゴリを理解内容のカテゴリと決定する。

ここでは、全カテゴリパターンのカテゴリスコアの中から最も高いカテゴリスコアを持つカテゴリパターンをカテゴリ理解結果とする。

3.3.3. 単語理解

単語理解では、カテゴリ理解結果をもとに、対話履歴の中から最もスコアの和が高い単語の組み合わせを、言語理解内容とする。

4. 評価実験

本稿で示してきた言語理解手法と音声認識結果の第一候補を最優先する言語理解手法、クラススコアを用いたカテゴリ理解を行う言語理解手法[5]の性能を比較するために評価実験を行った。実験を行うために、最新の認識結果のn-bestの第1候補を最優先するシステム(以下、SYS-a)、クラススコアを用いたカテゴリ理解を行う言語理解手法を採用したシステム(以下、SYS-b)、本稿で示した言語理解手法を採用したシステム(以下、SYS-c)を作成した。3つのシステムは言語理解部以外の性能、SYS-bとSYS-cは言語理解部のカテゴリ理解以外の性能は全て同等である。ただし、SYS-aではカテゴリ理解は行っていない。

予備実験として、情報系学部生・大学院生 15名にあらかじめ用意していた浜松西インターを目的地に設定する際に使用する21種類の発話(『静岡県』や『東

名自動車道』、『静岡県の東名自動車道』など)を1人3回ずつ発話してもらい、音声認識を行った。そこで得られた発話データに車内雑音データAを加えた音声データから第1発話-システム応答-第2発話といった形式(以下、U1-S1-U2)の対話データ(対話データA)を、車内雑音データAよりノイズが大きい車内雑音データBを加えた音声データからU1-S1-U2対話データ(対話データB)を生成した。この両対話データを目的地設定タスクに用いる対話データとした。また車内機器操作タスクに用いる対話データは[6]で用いられた発話データ(12種類の発話を20人に発話してもらったデータ)からU1-S1-U2対話データ(対話データC)として作成した。それぞれの対話データに用いた音声データの単語認識率は対話データAから順に71.46%、58.21%、50.60%であった。対話データA,B,CをそれぞれSYS-a,SYS-b,SYS-cに与え、U2発話に対するシステムの言語理解内容(以下、理解内容)と実際にユーザが発話した内容(以下、正解内容)を比較して、1対話当りでのカテゴリ理解精度、クラス単位のキーワード組の完全一致率、端誤理解精度を求めた。カテゴリ理解精度とは理解内容のカテゴリと正解内容のカテゴリが一致している割合を、完全一致率とは理解内容と正解内容の各クラス単位での単語が一致している割合を、単語理解精度とは理解内容の単語と正解内容の単語の正誤関係を示したものである。完全一致率と単語理解精度を総称して言語理解精度と呼ぶ。

4.1. カテゴリ理解精度の比較

SYS-bとSYS-cのカテゴリ理解精度の比較を行った結果を表1に示す。カテゴリ理解精度とは、理解内容のカテゴリが正解内容のカテゴリと一致している割合である。理解内容のカテゴリと正解内容のカテゴリが一致していないと完全一致になり得ないので、カテゴリ理解精度は完全一致率の上限となる。すなわち、カテゴリ理解精度は完全一致率(言語理解精度)において重要な処理であることが言える。

結果を見てみると、SYS-cは対話データA,Bの詳細化や回答で多少SYS-bに劣る結果となったが、その他の発話タイプや全体での精度を見てみると、SYS-bよりもカテゴリ理解精度が上がるのが分かった。このことから、クラススコアを用いたカテゴリ理解を行う言語理解手法(以下、SYS-bの言語理解手法)よりも提案手法の方が、カテゴリ理解精度が高いことが証明された。

4.2. 言語理解精度の比較

3つのシステムの言語理解精度の比較を行った結果を表2に示す。

まず完全一致率について比較する。対話データA,B,Cいずれにおいても、カテゴリ理解を行わない

表 1: 対話データ A, B, C におけるカテゴリ理解精度

対話データA: 31320対話、対話データB: 29490対話、対話データC: 7336対話						
システム	対話データ	全体(%)	詳細化(%)	訂正(%)	回答(%)	再入力(%)
SYS-b	A	28688(91.60)	25381(94.95)	655(51.98)	2483(81.14)	169(62.59)
SYS-c	A	29246(93.38)	25359(94.87)	979(77.70)	2653(86.70)	255(94.44)
SYS-b	B	23535(79.81)	16365(91.84)	959(32.29)	5855(75.94)	356(35.96)
SYS-c	B	24717(83.81)	16215(90.99)	1821(61.31)	5770(74.84)	911(92.02)
SYS-b	C	4780(65.16)	2358(51.53)	1661(88.97)	635(91.76)	126(62.69)
SYS-c	C	6399(87.23)	3805(83.11)	1756(94.05)	674(97.40)	166(82.59)

表 2: 対話データ A, B, C における各システムの言語理解精度

対話データA: 76500単語、対話データB: 69480単語、対話データC: 19248単語							
システム	対話データ	完全一致(%)	正解(%)	置換(%)	挿入(%)	脱落(%)	単語正解精度(%)
SYS-a	A	27477(87.73)	74343(97.18)	1699(2.22)	2804(3.67)	458(0.6)	93.52
SYS-b	A	28244(90.18)	74599(97.52)	1100(1.44)	2048(2.68)	801(1.05)	94.84
SYS-c	A	29067(92.81)	74645(97.58)	639(0.84)	1043(1.36)	1216(1.59)	96.21
SYS-a	B	20692(70.09)	63902(91.97)	4971(7.15)	6402(9.21)	607(0.87)	82.76
SYS-b	B	22072(74.77)	64451(92.76)	4118(5.93)	5082(7.31)	911(1.31)	85.45
SYS-c	B	24048(81.46)	64429(92.73)	2794(4.02)	2608(3.75)	2257(3.25)	88.98
SYS-a	C	1965(26.79)	9757(50.69)	2068(10.74)	79(0.41)	7423(38.57)	50.28
SYS-b	C	2682(36.56)	10223(53.11)	6169(32.05)	72(0.37)	2856(14.84)	52.74
SYS-c	C	3501(47.72)	13789(71.64)	4580(23.79)	122(0.63)	879(4.57)	71.00

(注: 完全一致率以外は正解内容に含まれる単語当たりの割合)

SYS-a よりカテゴリ理解を行う SYS-b, c の方が、完全一致率が高い結果となった。よって、音声認識結果の第 1 候補を最優先する言語理解手法よりもカテゴリ理解を行う言語理解手法の方が、完全一致率が高くなることが分かった。また、SYS-b の言語理解手法より提案手法の方が、完全一致率が高い。これは、SYS-b の言語理解手法より提案手法の方が、カテゴリ理解精度が高いことが理由として考えられる。

次に単語理解精度について比較する。対話データ A, B において脱落の割合が SYS-a よりも SYS-c の方が大きくなっていることが分かる。脱落が増える原因としては、カテゴリ理解結果に沿った単語の組み合わせが存在しなかったことが考えられる。また、対話データ C において、置換の割合が SYS-a よりも SYS-c の方が大きくなっている。置換が増える原因としては、カテゴリを正しく理解したが、単語を正しく理解できていないことが考えられる。両原因を解決するためには単語スコア生成の精度を向上させることが考えられるが、本提案手法ではカテゴリ理解精度の向上による言語理解精度の向上を目的としてきたので単語スコア生成については手を加えていない。今後、単語理解精度の向上のためにも、単語スコア生成についての検討が必要である。

5. まとめ

本稿では、音声認識結果の信頼度と対話履歴を利用した発話意図推定手法を用いた言語理解手法を示し、n-best の第 1 候補を最優先する言語理解手法とクラススコアを用いたカテゴリ理解を行う言語理解手法との評価実験を行った。評価実験により、n-best の第 1 候補を最優先する言語理解手法やクラススコアを用いたカテゴリ理解を行う言語理解手法よりも音声認識結果の信頼度と対話履歴を利用した発話意図推定手法を用

いた言語理解手法の方がカテゴリ理解精度と単語理解精度(言語理解精度)が高くなることが分かった。また、単語スコア生成における単語スコアの精度向上によって、言語理解精度が向上することも分かった。今後の課題としては、単語スコア生成の枠組みについて検討し、言語理解精度の向上を図ることが挙げられる。

参考文献

- [1] 甲斐充彦、石丸明子、伊藤敏彦、小西達裕、伊東幸宏 “目的地設定タスクにおける訂正発話の特徴分析と検出への応用” 日本音響学会全国大会論文集、2-1-8, pp. 63-64, 2001
- [2] 平沢純一、宮沢昇、相川清明 “質問-応答連鎖からの音声対話システムの誤解の検出”、情報処理学会研究報告、SLP-34-41, pp. 239-244, 2000
- [3] 駒谷和範、河原達也 “音声対話システムにおける音声認識結果の信頼度の利用法” 日本音響学会講演論文集、3-5-2, pp. 73-74, 2000
- [4] 駒谷和範、河原達也 “音声認識結果の信頼度を用いた頑健な混合主導対話の実現方” 情報処理学会研究報告、SLP-30, pp. 39-44, 2000
- [5] 水谷誠、伊藤敏彦、甲斐充彦、小西達裕、伊東幸宏 “音声認識の信頼度と対話履歴を利用した最尤推定型言語理解” 情報処理学会研究報告、SLP-45, pp. 113-118, 2003
- [6] 由浅裕規、水野智士、伊藤敏彦、甲斐充彦、小西達裕、伊東幸宏 “状況と文脈を利用した音声対話型車載インタフェースの構築と評価” 電子情報通信学会技術研究報告、SP2003-113~153, pp. 199-204, 2003
- [7] 中川聖一、甲斐充彦 “文脈自由文法制御による One Pass 型 HMM 連続音声認識法” 電子通信学会論文誌、Vol. J76-D-II, No. 7, pp. 1337-1345, 1993