6P-06

SVM と CRF を用いたロボットによるロバストな命令理解

小堀　嵩博 †　　中村　友昭 †　　長井　隆行 †
中野　幹生 ‡　　船越　孝太郎 ‡　　金子　正秀 †
† 電気通信大学　　‡（株）ホンダ・リサーチ・インスティチュート・ジャパン

1　はじめに

一般にロボットにおいては予めプログラミングされた行動を実行することが想定されており，その行動を変更する場合にはプログラムを変更する必要があった．しかし，一般のユーザが必ずしもプログラムの知識を持っているとは限らず，特に家庭用ロボットにおいてユーザがプログラムを変更することは困難である．本稿では，音声による自然な命令に対してロボットがロバストな命令理解を行う方法について述べる．ここでの自然な命令とは人が普段話す言語であり，同じ内容の命令であっても多様な言い回しが存在するため，柔軟に理解する必要がある．そのような柔軟な自然言語理解は，様々なタスクが要求される家庭用ロボットにおいて重要である．

近年，ロボットによる音声命令の理解への期待が高まっており，家庭用ロボットの性能を競う大会 Robocup@home において，音声命令の理解性能を競う課題 General Purpose Service Robot(GPSR) が行われている [1]．しかし，未だ高得点を取れるチームが少なく非常に難しい課題となっている．

本稿では，この GPSR タスクを対象とし，Support Vector Machine(SVM) による行動識別と Conditional Random Field(CRF) による名詞抽出 (スロット抽出) を組み合わせ自然言語理解を行う．図 1 に提案手法の概要を示す．まず，Julius[2] により音声命令の認識結果として n-best が得られ，それぞれの命令文を Mecab[3] により形態素解析する．その表層形，原形，品詞をもとに CRF によるスロット (物体名，人名等) 抽出を行う．また，命令文は単語分割され Bag Of Words(BOW) 表現にし，単語と行動との関係を学習した SVM により行動識別を行う．行動とスロットはそれぞれ n-best の解析結果が出力されるため，行動とスロットの共起関係を考慮し，尤もらしい組み合わせを選択する．以上のように，音声認識，行動識別，スロット抽出，行動とスロットの共起関係を統合することで，ロボットによるロバストな言語理解を行う．

従来，音声命令の理解が可能なロボットは存在したが，多くが特定のコマンドや言い方でないと理解できず，命令の方法を覚える必要があった．また，構文解析による命令理解 [4] も存在するが，音声認識結果が誤っている場合や，類義語辞書に登録されていない単語が命令文に存在すると行動を認識できず，加えてスロット抽出のルールを人が作る必要があった．一方，提案手法は機械学習により自動的にルールを学習することができ，音声認識誤りや学習辞書に存在しない単語が多少あったとしても，汎化性能により命令を理解することができる．

2　命令文データセット

本稿では GPSR において使用される命令を自動生成する GPSR 文生成器 [5] を使用し，連続した 3 つの内容の命令からなる命令文を収集した．また，より自然な命令を集めるため，Robocup@home に参加したことのある学生にアンケートを取り，1 文のみで構成された命令文を収集した．

Robust Instruction Understanding by Robot using SVM and CRF
†Takahiro KOBORI　†Tomoaki NAKAMURA　†Takayuki NAGAI
‡Mikio NAKANO　‡Kotarou FUNAKOSHI　†Masahide KANEKO
†The University of Electro-Communications
‡Honda Research Institute Japan Co., Ltd

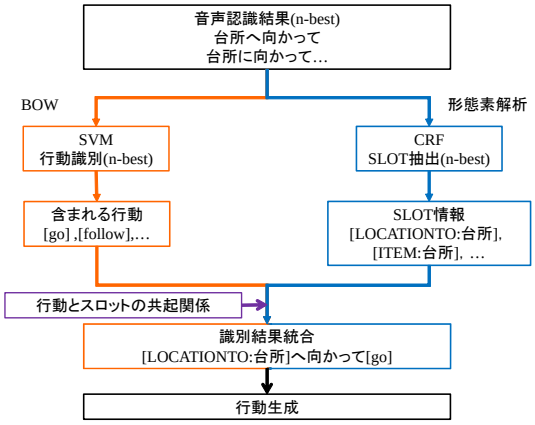

図 1　提案手法の概要

表 1　行動の種類

物体を運ぶ	物体を持ってくる
物体を把持する	物体を探す
目的地へ向かう	部屋から出る
初期位置へ戻る	自己紹介する
人についていく	人を探す
人を覚える	人を認識する
人に物を手渡す	

集められた命令文に対してスロット (物体名，人名等) 部分のアノテーションを行い，スロット部分を様々な単語に置換することで，内容の違う命令文をより多く生成した．以下が，収集した命令文の一例である．

- キッチンに行って，ジュースを取って，戻ってきて
- サイドテーブルからリモコンを持って来て
- ・・・

このような命令文を 1569 文集めた．スロットの種類は，物体名，物体の存在場所，目的地，人名の 4 種類であり，行動の種類は，表 1 の 13 種類である．

3　提案手法

3.1　音声認識

音声認識結果としては n-best が出力され，解析結果 o_n とその尤度を得ることができる．しかし，音声認識の尤度は 1 位のものが非常に高い値となり，そのまま用いると，1 位の認識結果のみを利用することとなる．しかし，音声認識では 1 位ではなく下位の認識結果が正解の場合がある．そこで，下位の結果を利用できるように，音声認識スコア $S_{sr}(o_n)$ を尤度が高い順に 1.0，0.9，... と設定することで，極端に値が小さくならないようにした．

また，一度に連続で複数の命令がなされる場合があるため，CRF による文区切りの推定を行った．本稿のシステムでは，一文を一つの行動としており，文区切りの誤りが生じてしまうと，適切な行動識別ができない．そのため，高精度な文区切りの推定を行う必要がある．

3.2　行動とスロットの共起関係に基づくスコア

SVM，CRF の結果はそれぞれ m-best，l-best 出力され，ある命令文 o_n に対するそれぞれの解析結果 $a_m(o_n)$,

$s_l(o_n)$ とスコア $S_{svm}(a_m(o_n))$，$S_{crf}(s_l(o_n))$ を得ることができる．この結果を統合するため，行動とスロットとの結びつきを表現した以下のようなスコア $S_f(a_m(o_n), s_l(o_n))$ を導入した．

$$S_f = \frac{2 \times (Recall) \times (Precision)}{(Recall) + (Precision)} + \phi \quad (1)$$

$$Recall = \frac{(s_l(o_n) \text{ のうち行動 } a_m(o_n) \text{ で必要なスロット数})}{(\text{行動 } a_m(o_n) \text{ で必要なスロット数})} \quad (2)$$

$$Precision = \frac{(s_l(o_n) \text{ に含まれる行動 } a_m(o_n) \text{ で必要なスロット数})}{(s_l(o_n) \text{ のスロット数})} \quad (3)$$

このスコアは精度 (Recall) と再現率 (Precision) との調和平均である F 値の考えを利用し，行動に必要なスロットを評価するだけではなく，スロットが過剰に抽出された場合，その値が減少する．すなわち，このスコアは行動に必要なスロットがどれだけ抽出されたかを示している．ここで ϕ はスロットが 1 つも抽出されなかった場合であっても，スコアが 0 となることを防いでいる．

3.3 識別結果の統合

最終的に，音声認識結果の n-best を考慮し，SVM と CRF それぞれのスコアを統合したスコアを計算することで，行動生成を行う行動とスロットを決定する．すなわち，以下の式を最大とする行動 $a_m(\hat{o_n})$，スロット $s_l(\hat{o_n})$，音声認識結果 $\hat{o_n}$ を決定する．

$$a_m(\hat{o_n}), s_l(\hat{o_n}), \hat{o_n} = \underset{a_m(o_n), s_l(o_n), o_n}{\arg\max} S_{svm}(a_m(o_n))^{\alpha} S_{crf}(s_l(o_n))^{\beta} S_f(a_m(o_n), s_l(o_n))^{\gamma} S_{sr}(o_n)^{\delta} \quad (4)$$

ただし，$\alpha, \beta, \gamma, \delta$ はそれぞれのスコアに対する重みパラメータであり，実験により決定する．

4 実験

4.1 交差検定による評価

2 章で集めた命令文データを使用し 10 分割の交差検定を行った．1 回の解析で 1569 文中 9 割を SVM, CRF の学習用データに使用し，残り 1 割を認識させ，正しく行動識別とスロット抽出ができるか評価した．今回の交差検定では音声認識誤りのない命令文に対する命令の理解性能を評価するため，式 (4) から音声認識を除く，$S_{svm}(a_m)$，$S_{crf}(s_l)$，$S_f(a_m, s_l)$ の積を用いた手法 (提案手法 1) で評価した．すなわち以下の式を最大とする行動 $\hat{a_m}$ と，スロット $\hat{s_l}$ を抽出する．

$$\hat{a_m}, \hat{s_l} = \underset{a_m, s_l}{\arg\max} S_{svm}(a_m)^{\alpha} S_{crf}(s_l)^{\beta} S_f(a_m, s_l)^{\gamma} \quad (5)$$

比較として，以下のように SVM, CRF のスコアが最大の 1-best のみを用いた手法を使用した．

$$\hat{a_m} = \underset{a_m}{\arg\max} S_{svm}(a_m) \quad (6)$$

$$\hat{s_l} = \underset{s_l}{\arg\max} S_{crf}(s_l) \quad (7)$$

表 2 が結果であり，1569 文中行動とスロットが正しく認識できた割合を正答率として示している．提案手法の方がより正しく理解できた文章が増えており，ロボットの命令の理解性能を向上することができた．

表 2　交差検定による命令理解精度

	認識正解文章	正答率
1-best	1445/1569	92.1%
提案手法	1467/1569	93.5%

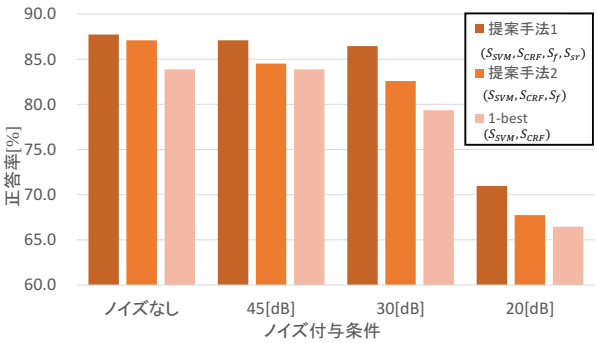

図 2　音声認識による命令理解精度

4.2 音声認識による命令理解

次に，2 章で集めたデータの中からランダムに 155 文を選び出し，式 (4) を用いた音声認識も含めた手法 (提案手法 2) の評価を行った．SVM，CRF に用いられる学習用データは残りの 1414 文であり，重みの設定は $\alpha = 1.9$，$\beta = 0.4$，$\gamma = 1.1$，$\delta = 0.8$ とした．比較として，4.1 節で用いた 1-best のものと，音声認識スコアを除いた提案手法 1 を使用した．提案手法 1 の重みは，$\alpha = 1.1$，$\beta = 0.5$，$\gamma = 2.1$ とした．4.1 節と同様に，155 文中正しく認識できた割合を正答率とする．また，ノイズの影響を比較するため，ノイズを付与しないものとホワイトガウスノイズを付与したものを比較した．ノイズは SN 比 45[dB]，30[dB], 20[dB], のものを付与した．図 2 が結果である．行動とスロットとの共起関係を示すスコア $S_f(a_m(o_n), s_l(o_n))$，音声認識のスコア $S_{sr}(o_n)$ を加えていくことで，徐々に正答率を上げることが出来ており，1-best と比較し，高い精度で命令文を認識することができた．また，ノイズが付与された環境でも，提案手法が最も高い正答率を示しており，音声認識誤りにもロバストな言語理解が可能である．

5 まとめ

本稿では，機械学習による言語理解と，行動とスロットの共起関係，音声認識結果を統合した言語理解手法を提案し，Robocup@home での General Perpose Service Robot タスクへ適用した．SVM と CRF を独立に用いるのではなく，行動とスロットの共起関係，音声認識結果を統合することで，より高精度かつノイズにロバストな言語理解が可能になることを示した．今後，ロボットによる知覚情報や文脈情報を本手法と統合することで，より高精度な言語理解の実現を図る予定である．

参考文献

[1] “Robocup@home,”http://www.ai.rug.nl/robocupathome/.
[2] “Julius,” http://julius.sourceforge.jp/.
[3] “Mecab: Yet Another Part-of-Speech and Morphological Analyzer,” https://code.google.com/p/mecab/.
[4] 板谷純希, 中村友昭, 長井隆行, “ユーザとのインタラクションに基づく学習を利用したロボットのタスクプログラミング,” The 25th Annual Conference of the Japanese Society for Artifical Intelligence, 3B1-OS22c-9, 2011.
[5] “RoboCup@Home 2011 General Purpose Service Robots 文生成器，” http://komeisugiura.jp/software/software_jp.html.