

## 音声対話システムにおける発話意図推定

松本 宗也† 樽松明† 白井克彦†  
†早稲田大学理工学術院

## 1 はじめに

音声対話システムにおいて、ユーザの発話意図に対して適切なインタラクションを行う事は重要である。発話意図を推定する手法としてはルールベースによる手法とコーパスベースの手法があり、コーパスを利用した手法では、統計ベースの推定方法 [1] や、事例ベースの方法が提案されている [2, 3]。しかし、これまでは形態素と過去の発話における発話意図（発話意図系列）から発話意図の推定を行っており、高い精度で発話意図を推定するにはより多くの情報が必要であると考えた。本稿では形態素や発話意図系列の他に文末特徴や形態素の n-gram を特徴要素としてサポートベクターマシン (Support Vector Machine, SVM) を利用して発話意図の推定を行い、どの特徴が推定に有効であるのかを示した。

## 2 SVM による発話意図推定

SVM は識別器の中で認識性能の優れた学習モデルの一つであり [4]、マージン最大化やソフトマージン、カーネルトリックにより未学習データに対してもデータの特徴量の次元が大きくなっても高い識別性能を得る事ができる。そのため発話のような多くの情報を利用する場面において SVM が有効であると考えた (図 1)。

SVM による発話意図推定手法は次に示す 1~4 のタグ付与済みコーパスによる学習と 5 の発話意図推定の 2 つの段階で構成され、次のような流れになっている。

1. タグ付与したコーパスを形態素解析器、係り受け解析器を用いて発話文の形態素解析・係り受け解析を行う
2. 1 の結果及び発話意図系列の統計を取得する
3. 前発話との類似度・形態素包含度と 1, 2 で取得した各情報と併せて特徴ベクトルとして定義する
4. タグ付与済みコーパスを解析し、定義した特徴について SVM で機械学習する
5. 発話文を解析し、4 で学習したモデルにより発話意図の推定を行う

発話意図の推定に利用する特徴量としては、前発話との類似度、発話と前発話における形態素包含度、発話意図系列 (過去の連続 3 発話まで)、形態素、文頭形態素、文末形態素、形態素の 2-gram 及び 3-gram を利用している。前発話との類似度は形態素の一致度から計算し、入力発話  $U_x$  と事例発話  $U_y$  の類似度  $SIM(U_x, U_y)$  は次のような式で求められる。

$$SIM(U_x, U_y) = \frac{2M_{xy}}{M_x + M_y} \quad (1)$$

ここで、 $M_x$ ,  $M_y$  はそれぞれ  $U_x$ ,  $U_y$  の形態素数、 $M_{xy}$  は一致する形態素数を表す。また、形態素包含度とは発

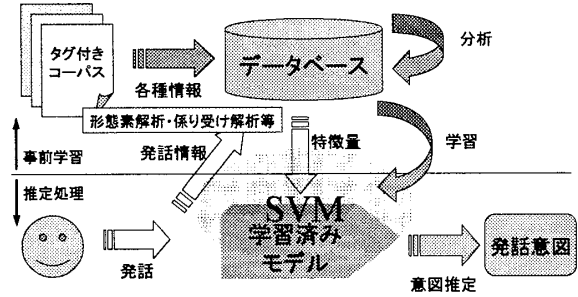


図 1: 発話意図推定の全体的な処理の流れ

話が前発話の形態素を含んでいる割合と前発話が発話の形態素を含んでいる割合である。形態素包含度はそれぞれ  $M_{xy}/M_x$ ,  $M_{xy}/M_y$  で求められる。特徴ベクトルの次元数は前発話との類似度が 1 次元、形態素包含度は 2 次元で固定であり、他の特徴量については 3 章で述べるコーパスから抽出したもので合計 56499 次元であった。

形態素解析、係り受け解析にはそれぞれ MeCab[5], Cabocha[6] を利用し、係り受け解析結果における文の最初の項を文頭、最後の項 (文全体を受けている項) を文末と定義し、文頭・文末に含まれる形態素等の情報を個別に特徴量として扱っている SVM には libsvm[7] を利用し、カーネルは radial basis function (rbf) を選択した。

## 3 発話単位タグ

本稿では荒木らによって提案された発話単位タグ [8] を発話意図として定義した。

コーパスは文部省重点領域研究「音声・言語・概念の統合的処理による対話の理解と生成に関する研究」音声対話コーパスワーキンググループ編集の音声対話コーパスの書き起こし、及び ATR 音声言語コミュニケーション研究所 音声データベースの「自然発話音声・言語データベース SDB TR1」の旅行会話から 4630 発話を加えたもので、計 5750 発話である。この時、ATR のコーパスの約半数に当たる 2334 発話に対して、Cohen の kappa 値を用いて評価を行った。Cohen の kappa 値 ( $\kappa$ ) は、観測された一致率を  $P(O)$ 、期待される一致率を  $P(E)$  とすると、次式で求める事ができる。

$$\kappa = \frac{P(O) - P(E)}{1 - P(E)} \quad (2)$$

追加したコーパスにおける評価は  $P(O) = 0.701$ ,  $P(E) = 0.122$ ,  $\kappa = 0.659$  であった。

Estimation of Utterance Intention of Spoken Dialogue System

Toshiya Matsumoto†, Akira Kurematsu†, Katsuhiko Shirait†  
†Faculty of Science and Engineering, Waseda University

表 1: 特徴量の変化と意図推定正解率 (%) の変化

推定に利用する特徴量	次元数	正解率
全て (形態素, 発話意図系列, 文頭形態素, 文末形態素, 類似度, 包含度, 2-gram, 3-gram)	56499	69.34
形態素	2589	60.64
形態素+文末形態素	3629	62.97
形態素+意図	3177	68.16
形態素+意図+文頭	4929	67.79
形態素+意図+2-gram, 3-gram	14068	69.01
形態素+意図+文末	5074	70.30
形態素+意図+文末+類似度	5075	70.28
形態素+意図+文末+包含度	5076	70.43
文頭以外全て	52448	69.53
文頭以外 (2-gram, 3-gram 調整)	11868	70.59

## 4 発話意図推定手法の評価実験

前述のコーパスを用いて発話意図推定実験を行った。正解率の算出には全ての実験において、コーパスを12に分割した12-fold cross validationで行った。

### 4.1 特徴量変化による発話意図推定

SVMを用いた意図推定手法において特徴量と発話意図推定の関係を調べるため、特徴量を変化させて正解率を比較する評価実験を行った(表1)。

実験の結果から、特に発話意図系列や文末特徴が発話意図の推定に有効であり、形態素包含度を利用する事で僅かではあるが精度を向上できる事が確認できた。これは、前発話の内容を含んだ発話には意図が偏りやすい傾向があるからであると考えられる。一方、前発話との類似度や文頭形態素、形態素2-gramや3-gramは特徴を加える事で精度を落とす結果となった。文頭特徴に関しては、フィラーや相槌のような発話が含まれやすい事が原因であると考えられる。2-gramや3-gramに関してはそれだけで次元数が非常に大きくなってしまふのが原因と考え、2-gramと3-gramにおいて出現数による閾値で特徴次元を圧縮してから特徴量として加えた結果、精度を向上する事が可能であった。形態素の他にも文末といったような文構造や前発話との関係に関する情報も発話意図の推定には有効である事が分かった。また、2-gramや3-gramのように特徴次元が大きくなりすぎてしまう情報を扱う場合はその一部を推定に利用する事で精度が向上でき、逆に文頭特徴や前発話との類似度のように発話意図の推定に利用できない特徴が存在する事も確認できた。

### 4.2 発話意図推定の比較実験

事例に基づく発話意図推定において、コーパス中の各発話との類似度を計算することで発話意図を推定する手法[3]と本稿のSVMによる手法で発話意図推定の比較実験を行った(表2)。ここで、各発話との類似度の計算方法として正解率が高かった形態素の一致度を採用した(式1)。比較実験ではコーパス全てを利用したものとATRのコーパスの2種類を用いた。

表 2: SVM, 形態素のみ利用したSVM, 事例に基づく発話意図推定手法の比較 (正解率: %)

	SVM	SVM (形態素)	事例
コーパス全て	70.59	60.64	50.64
ATR	72.73	64.99	51.98
学習済み	99.57	95.27	93.91

比較実験の結果、SVMによる推定の方がいずれも事例に基づく発話意図推定より良い結果となった。今回、事例に基づく手法の推定手法の正解率は50%強と低い結果となったが、これは使用したコーパスに様々な話題が含まれていたため、学習したコーパス中に発話文と類似した事例が少なかったのが理由であると考えられる。ATRのコーパスで同様の実験を行うと、いずれの手法でも精度の向上が見られたが、これは学習したコーパスと推定を行う発話の内容が近くなったためと考えられる。また、学習済みのコーパスにおいては非常に高い精度で発話意図の推定が行える事から学習するコーパス量や内容を変化させる事でより高い精度での推定が可能である事が分かる。

## 5 まとめ

本稿では発話に含まれる様々な特徴からSVMにより発話意図推定を行い、どのような特徴が発話意図の推定に有効であるかを検討した。実験の結果から、発話意図の推定に利用できる有効な特徴として、形態素や発話意図系列の他に文末特徴、前発話との形態素包含度、調整した形態素n-gramが確認できた。単語のソーラス情報など発話意図の推定にはまだ利用されていない有効な特徴が存在する事が考えられ、それら多くの情報を利用することでより高精度な発話意図推定の実現が期待できる。

## 謝辞

本研究の一部は、早稲田大学理工学研究所の研究課題「自発的コミュニケーション機構を有するマルチモーダルヒューマンインタフェースの研究」、平成19年度科学研究費基盤研究(B)課題番号17300066「対話状況に応じた自発的コミュニケーション機構の研究」によるものである。

## 参考文献

- [1] 白木将幸, 伊藤敏彦, 甲斐充彦, 中谷広正, “自然発話文における統計的な意図理解手法の検討,” 情報処理学会研究報告.SLP, 音声言語情報処理, 2004-SLP-50-11, Feb. 2004.
- [2] 入江友紀, 松原茂樹, 河口信夫, 山口由紀子, 稲垣康善, “意図タグつきコーパスを用いた発話意図推定手法,” 人工知能研究会資料, SIG-SLUD-A301-03, Jul. 2003.
- [3] 岩下薫, 嶋田和孝, 遠藤勉, “対話システムにおけるユーザの発話意図推定,” 火の国シンポジウム2007 CD-ROM, A-3-2, Mar. 2007.
- [4] Nello Cristianini, John Shawe-Taylor 著, 大北剛訳, “サポートベクターマシン入門,” 共立出版, Mar. 2005.
- [5] 工藤 拓, MeCab/和布蕪, <http://mecab.sourceforge.jp/>
- [6] 工藤 拓, CaboCha/南瓜, <http://www.chasen.org/~taku/software/cabocha/>
- [7] Chin-Chung Chang, LIBSVM, <http://www.csie.ntu.edu.tw/~cjlin/libsvm/>
- [8] 荒木雅弘, 伊藤敏彦, 熊谷智子, 石崎雅人, “発話単位タグ標準化案の作成,” 人工知能学会誌, Vo.14, No.2, Mar. 1999.