No News規準を用いた韻律情報の意味学習

The meaning acquisition of prosodic information by No News Criterion

田中 一晶[1*]　中谷 仁[1]　高田 宏明[2]　冨永 善視[2]
TANAKA Kazuaki[1], NAKATANI Hitoshi[1], TAKADA Hiroaki[2], TOMINAGA Yoshimi[2],
尾関 基行[1]　岡 夏樹[1]
OZEKI Motoyuki [1], OKA Natsuki[1]

[1] 京都工芸繊維大学 大学院工芸科学研究科
[1] Graduate School of Science and Technology, Kyoto Institute of Technology
[2] 京都工芸繊維大学 工芸科学部
[2] School of Science and Technology, Kyoto Institute of Technology

Abstract: In human-robot interaction, it is very important for robots to understand positive and negative evaluations given by humans. Most of previous works focused on prosodic information and classified positive and negative evaluations with high accuracy. However, these works used supervised learning, and so they need hand labeling and/or prior knowledge. In this work, we also focus on prosodic information, and propose a method that classifies prosodic information into positive and negative evaluation by unsupervised learning based on No News Criterion, which interprets the lack of utterance as a positive evaluation.

1　はじめに

人ーロボットインタラクションでは、ロボットが人からの肯定・否定の評価を理解できることは、大変重要である。この理解に基づいてロボットは、不適切だった行動を修復すべく行動し直すことができるし、さらに、この評価に基づいて学習しておくことによって、将来の行動をより適切なものにすることができる。これまでに、韻律情報を用いて肯定・否定や人の感情状態を識別する研究が多く行われてきた [1, 2, 3, 4]。

[2] では、韻律情報を複数の意味に分類し、特に肯定・否定、聞き返し、フィラーといった発話の識別に韻律情報が適していることを示している。[1] では、終助詞「ね」の意味が賞賛であるか皮肉であるかを韻律情報に基づいて識別している。また、韻律情報と非言語情報を併用することで人とのコミュニケーションの円滑化を図る研究も行われている [3, 4]。[3] では、視線と韻律情報から人の感情を推測し、それに応じてロボットが感情表出を行うことで人との自然な会話を実現している。[4] では、韻律情報と頭部ジェスチャを併用することで肯定・否定の度合いを識別できる手法を提案し、効率的な対話が実現できることを示している。

*連絡先：京都工芸繊維大学 大学院工芸科学研究科
〒 606-8585 京都市左京区松ヶ崎橋上町
E-mail:d8821007@edu.kit.ac.jp

これらの研究では韻律情報とその意味との関連を高い精度で学習できているが、その学習には教師あり学習が用いられているため、実験者による手作業でのラベル付けや、事前知識を必要としている。これに対し、本研究では、一定時間発話が無いことを肯定的な評価と捉える No News 規準 (NNC)[5] を用いることで、教師なし学習で韻律情報の意味を肯定・否定に分類する方法を提案する。

2　インタラクションデータ収集実験

本節では、人とロボットのインタラクションデータを収集するために行った実験について説明する。ここで収集したデータを用いて韻律情報の意味学習 (3 節) を行う。

2.1　実験タスク

実験参加者 (大学生、大学院生の男性 5 名) には以下のゲーム (図 1 参照) を行ってもらった。

- 実験参加者は白い AIBO に言葉で指示 (「まえ」「うしろ」「ひだり」「みぎ」) を行い、部屋の中心に置かれた骨に白い AIBO を接触させると勝ち。

- 白い AIBO が実験者がコントローラにより遠隔操作する黒い AIBO に接触されると負け。
- 黒い AIBO は部屋の中央に位置する枠線内 ($1.5m \times 1.5n$ から出ることはできない。

実験参加者には以下の情報を予め伝えておいた。

- 指示に使用できる言葉は「まえ」「うしろ」「ひだり」「みぎ」の 4 種類である。
- AIBO は最初は言葉の意味が理解できておらず、指示通りに行動するとは限らない。
- 「よし」「だめ」の言葉を使って評価すると、AIBO は少しずつ言葉の意味を理解していく。
- 言葉を認識すると AIBO の耳が動く。

2.2 実験仕様

実験は図 1 のように室内に用意した $3.3m \times 3.3m$ の空間で行う。実験参加者の発話は連続音声認識ソフト Julius[6] を用いて認識するが、行動教示：「まえ」「うしろ」「ひだり」「みぎ」、評価教示：「よし」「だめ」を認識できるように、予めこれらの単語を登録しておく。また、それらの発話を認識すると、AIBO は耳を動かして認識できたことを示すと共に、前進・後退・左回転・右回転の行動と喜ぶ・悲しむの表出をそれぞれ行う。

行動教示を認識するとそれに対応する行動を確率 P で正しく実行し、$1-P$ の確率でそれ以外の行動をランダムに実行する。正しい行動の選択確率 P は、評価教示が与えられる度に以下の式によって更新する。

$$P_{t+1} = P_t + \alpha(P_{max} - P_t) \tag{1}$$

ここで、初期値 $P_0 = 0.25$，最大確率 $P_{max} = 0.75$，学習率 $\alpha = 0.04$ とする。

勝敗が決まれば 2 台の AIBO を定位置に戻し、ゲームを再開する。各実験参加者について 30 分程度実験を行った。実験開始から終了までの実験参加者の発話を録音し、発話ごとに、発話開始時刻、発話終了時刻、発話の認識結果、AIBO が実行した行動、正しい行動の選択確率、発話の遅れ時間 (AIBO が行動を開始してからその発話が与えられるまでの時間) をログデータとして記録した。ただし、発話の遅れ時間を記録するのは指示に対して AIBO が行動を実行した後の発話だけである。

2.3 肯定的発話と否定的発話

本研究では、ロボットが与えられた指示に対して正しい行動を実行したあとに与えられた発話を肯定的発話、正しい行動以外の行動を実行したあとに与えられた発話を否定的発話と呼ぶ。例えば、「まえ」という指示に対してロボットが後退した場合、人は否定的な評価や再度「まえ」という指示を与えるだろう。このときの発話が否定的発話である。また、「まえ」という指示に対して正しく前進した場合、人は肯定的な評価や次の指示を与えるだろう。このときの発話が肯定的発話である。

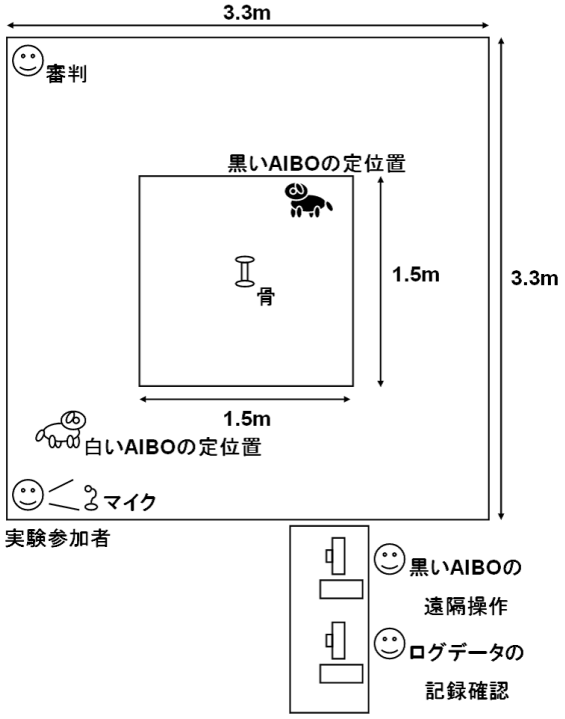

図 1: 実験環境

肯定的発話は遅れて与えられる傾向があるため [5]、この遅れ時間を肯定的発話の識別規準 (No News 規準) として利用する。また、我々は「即座に与えらる発話は否定的発話である」ことを新たに仮定し、否定的発話の識別にも遅れ時間を利用する[1]。

3 韻律情報の意味学習

我々が提案する韻律情報の意味学習は訓練例抽出フェイズ、韻律情報抽出フェイズ、否定的発話分類フェイズ、意味学習フェイズの 4 段階に分けて行う。本節では、各段階で行う処理について説明する。

3.1 訓練例抽出フェイズ

訓練例抽出フェイズでは、No News 規準 [5](一定時間、発話が無いことを肯定的な評価と捉える規準。以下、NNC) を用いて 2 節の実験から得られた発話のロ

[1]本研究で実施した実験 (2 節) では、即座に与えられる肯定的発話も多く確認されたため、否定的発話分類フェイズ (3.3 節) でさらに否定的発話の分類を行う。

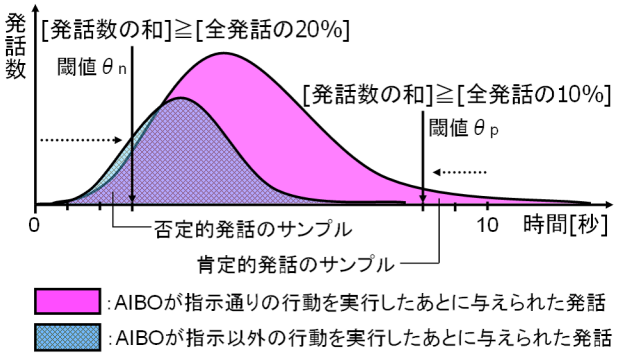

図 2: NNC による発話の分類手順
※横軸の時間は、(指示に対して)AIBO が行動を開始してからその発話が与えられるまでの時間である。

グデータを、発話の遅れ時間 (AIBO が行動を開始してからその発話が与えられるまでの時間) の長さで分類し、肯定的発話の訓練例と、否定的発話の訓練例を得る。

図 2 に NNC による発話の分類手順を示す。肯定的発話 (AIBO が指示通りの行動を実行した後に与えられた発話) は否定的発話 (AIBO が指示以外の行動を実行したあとに与えられた発話) よりも遅れて与えられることが多いため、発話の遅れ時間ごとにその回数を記録すると図 2 のような分布になる。肯定的発話の訓練例は閾値 θ_p を超える遅れ時間を持つ発話とする。θ_p は、最初 10 秒 に設定し θ_p を超える遅れ時間を持つ発話の数が全発話の 10% を超えるまで 0.1 秒間隔で短くしていくことで決定する。

否定的発話の訓練例は閾値 θ_n より短い遅れ時間を持つ発話とする。θ_n は、最初 0 秒 に設定し θ_n より短い遅れ時間を持つ発話の数が全発話の 20% を超えるまで 0.1 秒間隔で長くしていくことで決定する。しかし、図 2 のように否定的発話の訓練例には多くの肯定的発話が混ざっているため、否定的発話は後述の 3.3 節にてさらに高い精度で分類する。

また、本研究では、肯定的発話にも否定的発話にも分類されなかった発話を、後述の意味学習結果の評価 (4 節) に使用するテストデータとする。

3.2 韻律情報抽出フェイズ

ここでは、録音した発話から韻律情報を抽出する。抽出する韻律情報として、肯定と否定の発話を分類するのに有効とされている次の 7 つの韻律情報を採用した [1, 2, 3, 4, 7]。

- F0 値のピークピーク値
- F0 値の平均値
- F0 値の最大値
- F0 値の最小値
- F0 値の標準偏差

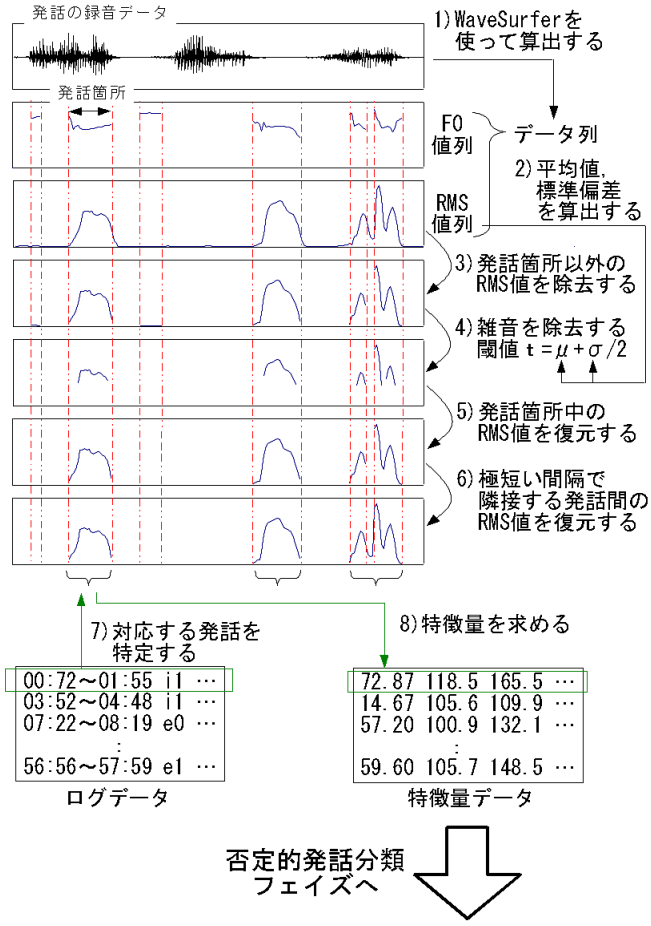

図 3: 韻律情報の抽出過程

- RMS 値の最大値
- RMS 値の最小値

韻律情報の抽出は次の手順で行う。まず、実験で得た発話音声列を音声分析ソフト WaveSurfer[2]にかけ、基本周波数 (F0 値) とパワー (音圧の RMS 値) のデータ列を算出する。そして、算出したデータ列のパワーの平均値に標準偏差の 1/2 を足し合わせて、閾値を作成する。発話箇所以外のパワーを除去し、次に作成した閾値を使って雑音データを除去する。雑音として誤って除去された発話箇所中のパワーを復元し、さらに、0.15 秒以下の極短い間隔で隣接する音声間のパワーを復元し、1 つの発話とみなして繋ぎ合わせる。最後に、抽出した韻律情報を発話時刻に基づいてログデータと対応付ける。

この韻律情報の抽出過程を図 3 に示す。

3.3 否定的発話分類フェイズ

ここでは、韻律情報抽出フェイズ (3.2 節) で得られた肯定的発話と (肯定的発話が多く混ざった) 否定的発話

[2]http://www.speech.kth.se/wavesurfer/

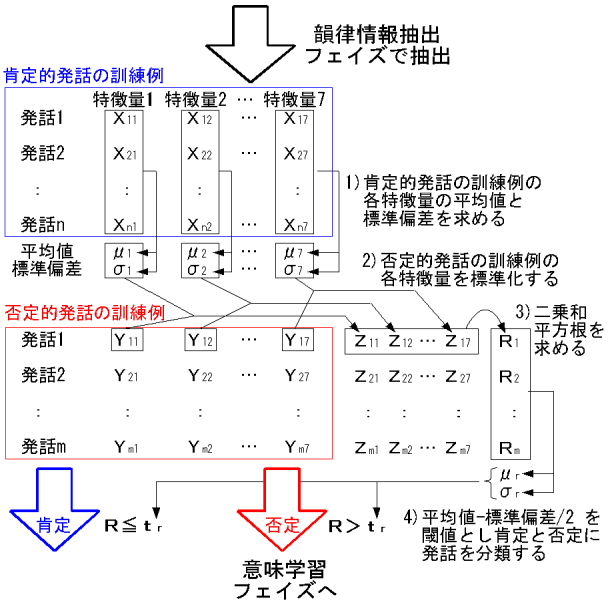

図 4: 否定的発話の分類

の韻律情報 (以下では韻律情報を特徴量と呼ぶ) から、より純度の高い否定的発話の特徴量を得る。

まず、否定的発話 (肯定的発話が多く混ざっている) の各特徴量を標準化し、その二乗和平方根を求める。次に、その二乗和平方根が、肯定的発話の特徴量の平均値と標準偏差から求めた閾値より大きければ否定的発話として、否定的発話の特徴量を得る。

詳しい手順を図 4 と以下に示す。

1. 肯定的発話の特徴量の平均 μ_1〜μ_7 と標準偏差 σ_1〜σ_7 を求める。
2. 否定的発話の特徴量 y_1〜y_7 を次の式で標準化する。$z = (y - \mu)/\sigma$
3. 否定的発話を各発話毎に標準化した特徴量群の二乗和平方根を求める。
4. 3. で各発話毎に求めた値の平均値 μ_r と標準偏差 σ_r 求め、閾値 t_r を作成し、閾値 t_r より大きければ否定的発話として分類する。

これにより、肯定的発話と否定的発話が混ざった特徴量の中から肯定的発話の特徴量に近いものを省くことで、より純度の高い否定的発話の特徴量を得ることができる。

3.4 意味学習フェイズ

意味学習フェイズでは、パターン認識において最も優秀な機械学習手法の一つであるとされている SVM (Support Vector Machine) を利用し、韻律情報の意味を学習する。学習に利用する特徴ベクトルは、韻律情報抽出フェイズで得られた肯定的発話の特徴量と否定的発話分類フェイズで得られた否定的発話の特徴量とする。

実装には TinySVM[3] を用いており、今回カーネル関数は RBF カーネル (2) と ANOVA カーネル (3) を用いて、試行錯誤的に結果が良いカーネルとパラメータを採用した。その結果、本研究においては ANOVA カーネルが適していることがわかった。用いたパラメータは 4 節に示す。

$$K(\mathbf{a},\mathbf{b}) = exp\left(-\frac{\|\mathbf{a}-\mathbf{b}\|^2}{\sigma^2}\right) \quad (2)$$

$$K(\mathbf{a},\mathbf{b}) = \left\{\sum_k exp\left(-s\|a_k - b_k\|^2\right)\right\}^d \quad (3)$$

4 意味学習の結果

我々が提案した韻律情報の意味学習方法を評価するため、意味学習フェイズ (3.4 節) において実験参加者 5 人それぞれのデータを用いて別々に学習を行った場合と、5 人分の学習データをひとまとめにして学習した場合、それぞれについて意味学習結果を評価した。その結果を表 1 の第 5〜7 列に示す。本研究で用いた SVM のカーネルは ANOVA カーネルであり、用いたパラメータは $s = 0.001$, $d = 1$ である。このパラメータは 5 人分の学習データを用いて学習した場合の正確度 (正しく識別された発話の割合) が高くなるように決定した。また、この評価に用いたテストデータは訓練例抽出フェイズ (3.1 節) で肯定的発話と否定的発話の訓練例として選ばれなかったものである。また、表の第 2〜4 列に訓練例抽出フェイズと否定的発話抽出フェイズでの、肯定的発話と否定的発話の分類精度 (肯定的 (否定的) 発話と分類された内、本当に肯定的 (否定的) 発話であるかを示す割合) を示す。ただし、否定的発話分類フェイズでは、否定的発話と分類された特徴量のみ学習データとして扱うため、否定的発話の分類精度のみを示す。

次に、AIBO が正しい (誤った) 行動を実行した回数と肯定 (否定) の評価が与えられた回数を表 2 に示す。

5 考察

テストデータを用いて行った意味学習結果の評価 (表 1 の第 5〜7 列) より、実験参加者ごとの正解度 (正しく識別された発話の割合) の差は大きいが、ある程度識別できていることがわかる。

まず、実験参加者ごとの正確度の差が大きい (90〜42%) 理由について考察する。否定的発話分類フェイズ

[3] http://chasen.org/~taku/software/TinySVM/

表 1: 訓練例抽出フェイズと否定的発話分類フェイズにおける肯定的発話・否定的発話の分類精度とテストデータを用いた意味学習結果の評価 (単位:%)

	訓練例抽出フェイズ		否定的発話分類フェイズ	テストデータを用いた意味学習結果の評価		
	肯定	否定	否定	肯定	否定	正解度
A	100	73	100	94	86	90
B	96	33	65	91	62	80
C	65	22	41	85	30	62
D	68	10	6	54	21	42
E	88	30	37	49	63	55
全体	84	77	41	73	49	64

※表中の A〜E は実験参加者であり、「全体」は 5 人分の学習データをひとまとめにして学習した場合の結果である。また、正確度とは肯定的発話と否定的発話両方が正しく識別された割合である。

表 2: AIBO が正しい (誤った) 行動を実行した回数と肯定 (否定) の評価が与えられた回数 (単位：回)

	AIBO の行動		実験参加者の評価	
	正しい行動	誤った行動	肯定	否定
A	81	54	28	54
B	157	77	131	75
C	137	75	121	73
D	158	62	142	58
E	104	52	73	42

※表中の A〜E は実験参加者である。

での分類精度 (表 1 の第 4 列) を見ると、最も正確度が悪かった 2 名の実験参加者 D，E の分類精度はそれぞれ 6%, 37% であり、他の実験参加者と比べて低く、高い正確度であった実験参加者 A，B の分類精度はそれぞれ 100%, 65% であり、否定的発話の分類精度が高いほど正確度が高いことがわかる。つまり、否定的発話の学習データを作成した段階 (否定的発話分類フェイズ) で実験参加者ごとの分類精度に大きな差があるため、それが正解度の差に影響していると考えられる。

否定的発話分類フェイズ (3.3 節) は、訓練例抽出フェイズ (3.1 節) で分類された (肯定的発話が多く混ざった) 否定的発話の訓練例から、肯定的発話の訓練例に近いものを除去し、より適切な否定的発話の訓練例を得るための段階である。肯定的発話の訓練例の分類精度 (表 1 の第 2 列) は、どの実験参加者もある程度高い (100〜65%) ため、これを基準に否定的発話の訓練例の分類を行うことに問題は無いと考える。実際、否定的発話分類フェイズでの肯定的発話の分類精度 (表 1 の第 4 列) は、訓練例抽出フェイズでの分類精度 (表 1 の第 3 列) と比べて (実験参加者 D 以外は)、30〜20%程度向上している。しかし、実験参加者 C, D, E の分類精度は 41〜6%であり相変わらず低いままである。これは、実験や AIBO の仕様によって即座に与えられた肯定の評価が多かったことが起因していると考えている。

本研究では、即座に与えられた発話は否定的発話であることを仮定しているが、「少しだけ前進して欲しい」ときに実験参加者が「まえ」という指示を AIBO に与えた場合、AIBO が正しく前進しても即座に次の指示を与えるだろう。また、表 2 を見ると実験参加者 A では AIBO が正しい行動を実行した回数 81 回に対し、評価を与えた回数は 28 回であり、他の実験参加者よりも少ないことがわかる。これは、AIBO が正しい行動を選択するようになるにつれて、実験参加者 A は肯定の評価を与えなくなっていったためである。他の実験参加者 (A 以外の 4 名) は毎回 AIBO の行動を評価しようとしていたため、直ぐに次の指示を与えたい情況においては、急いで評価を与え、続いて次の指示を与えるという場面が多く見られた。AIBO は徐々に正しい行動を実行しやすくなり、実験開始から 10 分前後経過すると 70%程度正しい行動を実行するようになるため、それに従って即座に与えられる肯定的発話も増加してしまう。本研究で実施した実験では、AIBO に評価を与えてもらうように実験参加者に説明を行ったため (2.1 節)、毎回 AIBO に評価を与えなければならないと解釈した実験参加者が多かったのであろう。しかし、人-ロボットインタラクションにおいて、ロボットが正しい行動を実行するようになるにつれて評価を与えなくなることは我々の先行研究でも多く観察されているため [8, 9]、そのような場合には実験参加者 A のように、肯定的発話だけでなく否定的発話においても高い分類精度が得られることが期待できる。

また、即座に与えられた肯定的発話は、遅れて与えられた肯定的発話とは異なる韻律情報を持っている可能性ある。我々は、そのために否定的発話分類フェイズにおいて、否定的発話の訓練例から除去しきれなかった肯定的発話があったのだと推測している。さらに、意味学習の評価に使用したテストデータは遅れて与えられた発話と即座に与えられた発話以外の発話であるため (詳しくは 3.1 節を参照)、発話のタイミングによって韻律情報が異なるのだとすると、評価方法の妥当性も検討する必要がある。

6 まとめと今後の展望

本研究では、一定時間発話が無いことを肯定的な評価と捉える No News 規準 (NNC) を用いることで、教師なし学習で韻律情報の意味を肯定・否定に分類する方法を提案した。また、即座に与えられた発話は否定的発話 (AIBO 正しい行動以外の行動を実行したあとに与えられる発話) であることを仮定し実験を行ったが、実験タスクやロボットの行動仕様によって肯定的発話も即座に与えられる場合があることがわかった。我々は、NNC によって高い精度で分類された肯定的発話の訓練例を元に、即座に与えられた発話の中から否定的

発話を取り出す段階：否定的発話分類フェイズ (3.3) を設けることで、即座に与えられた肯定的発話への対応を試みたが、即座に与えられた発話と遅れて与えられた発話とでは韻律情報が異なる可能性があるという問題が浮かび上がった。

我々は今後の取り組みとして、以下を予定している。

発話のタイミングと韻律情報の関係： 遅れて与えられた発話と即座に与えられた発話では異なる韻律情報を持つのか、実験で得られた発話データを分析し確認する。

SVM のカーネル関数・パラメータの自動決定： 韻律情報の意味学習には SVM を使用しているが、その識別精度に大きく影響を与えるカーネル関数や各種パラメータの選択は人が試行錯誤的に行っている。我々は韻律情報の意味学習を教師なし学習で行い、全てを自動化することを目標にしているため、適切なカーネル関数やパラメータを自動的に選択する方法を検討する。

自由発話での意味学習実験： 本研究では、発話データの収集を容易にするため、Julius に予め登録した語 (行動教示：「まえ」「うしろ」「ひだり」「みぎ」、評価教示：「よし」「だめ」) のみを実験参加者に使用してもらったが、我々の提案する方法は使用する語に制限の無い自由発話の意味学習にも適応できることを想定している。次の段階として、自然なインタラクションを通して得られた自由発話の意味学習実験を行う。

オンラインでの意味学習方法： 本研究では、韻律情報の意味学習はオフラインで行った。実用性を考えるとオンラインで学習できることが望ましいため、人から発話が与えられる度に逐次学習できる方法を検討する。

謝辞

本研究は科研費 (17500093 および 21500137) の助成を受けたものである。

参考文献

[1] 光本浩士, 濱崎敏幸, 大多和寛, 田村進一, 柳田益造：終助詞「ね」の韻律による皮肉と賞賛の識別, 電子情報通信学会論文誌, J84-D- No.5, pp. 851–853 (2001).

[2] 石井カルロス寿憲, 石黒浩, 萩田紀博：韻律および声質を表現した音響特徴と対話音声におけるパラ言語情報の知覚との関連, 情報処理学会論文誌, Vol. 47, No. 6, pp. 1782–1792 (2006).

[3] Breazeal, C. and Aryananda, L.: Recognition of Affective Communicative Intent in Robot-Directed Speech, *Autonomous Robots*, Vol. 12, No. 1, pp. 83–104 (2002).

[4] 藤江真也, 江尻康, 菊池英明, 小林哲則：肯定的/否定的発話態度の認識とその音声対話システムへの応用, 電子情報通信学会論文誌, J88-D- No.3, pp. 489–498 (2005).

[5] Tanaka, K., Zuo, X., Sagano, Y. and Oka, N.: Learning the meaning of action commands based on No News Is Good News Criterion, *Workshop on Multimodal Interfaces in Semantic Interaction*, pp. 9–16 (2007).

[6] 河原達也, 李晃伸：連続音声認識ソフトウエア Julius, 人工知能学会誌, Vol. 20, No. 1, pp. 41–49 (2005).

[7] 吉川哲史：機械学習を用いたノンバーバル発話の意図自動識別, 奈良先端科学技術大学院大学情報科学研究科, 修士論文, 62 pages (2000).

[8] 田中一晶, 岡夏樹：Scaffolding(足場づくり) を利用した学習系の構築, FIT2008 第 7 回情報科学技術フォーラム, RJ-006, 4 pages (2008).

[9] 田中一晶, 岡夏樹：人-ロボットインタラクションにおける「ためらう」ロボットの実験的評価, HAI シンポジウム 2008, 2B-2, 6 pages (2008).