3F3-6

音声行動コーパスに基づく多層常識推論モデルの構築

Multi-layered Commonsense Reasoning Model Construction based on Speech Behavior Corpus

桐山 伸也[*1] 大谷 尚史[*2] Heikki Ruuska[*3] 竹林 洋一[*4]
Shinya KIRIYAMA Naofumi OTANI Heikki RUUSKA Yoichi TAKEBAYASHI

[*1]静岡大学情報学部
Fuculty of Informatics, Shizuoka University

[*2]静岡大学大学院理工学研究科
Graduate School of Science and Engineering, Shizuoka University

[*3]静岡大学大学院情報学研究科
Graduate School of Informatics, Shizuoka University

[*4]静岡大学創造科学技術大学院
Graduate School of Science and Technology, Shizuoka University

We propose a methodology for studying commonsense reasoning frameworks based on the multi-level thinking model proposed by Marvin Minsky. The core method is "multimoda speech behavior corpus," which includes goal-oriented behavior description focusing speech. We describe corpus construction focusing infant whose behavior is naive and easier to observe thinking process than adult. We demonstrate how to design and verify commonsense reasoning models using the constructed corpus.

1. はじめに

音声コーパスの構築技術が進歩し，多言語化，マルチモーダル化により大規模，大容量化が進み，話し言葉の多様性，感情・意図を扱うなど深さ方向にも発展してきた（図 1）．日本語話し言葉コーパス [Furui 05] は，世界最大規模の自然発話音声研究用データベースであり，話し言葉の諸相の分析，話し言葉の音声認識・合成に幅広く活用されている．コーパスの収集対象は，固定した場所での会話や対話から，会議室や国会での発言，ロボットとの対話，クルマの中など実世界の現場へと拡大しており，国会議事録の自動作成など高度な話し言葉処理技術の開発が進んでいる [Kawahara 08]．音声関係の各種コーパスの着実な充実化により，音声認識や音声情報処理の研究実用化が進展した．しかしながら，人間の行動や思考のメカニズム解明に正面から取り組む研究例は少ない．特に，人間の常識を考慮に入れた音声コーパスの開発事例はない．人間の姿を真似たロボット開発やロボットのインタラクションの研究が盛んであるが [石黒 05]，人間らしく振舞っていても，人間の思考や感情とは全く異なるメカニズムで動作している．ロボットは 2 歳児の常識すら持ち合わせていないので，安定的・持続的な発展は望めない．長期的な基礎研究という視点で，コモンセンス知識とそれを用いた常識推論の研究が不可欠である．この観点から筆者らは，思考が行動に表出しやすいナイーブな幼児に注目し，マルチモーダル幼児教室を基盤として，実世界の行動観測に基づく感情・意図・思考の多層記述を持つ音声行動コーパスを構築している．本稿では，音声行動コーパスに基づく常識推論モデル構築の方法論について述べる．

2. マルチモーダル音声行動コーパス

人間の感情・意図・思考を表出するメディアとして，音声は一次元情報として観測可能で文字シンボルに書き起こせるので，画像や他のセンサ情報よりも扱いやすい．この観点から，音声を機軸に感情・意図・思考などの内面的特徴に踏み込んで多層的に行動を記述するマルチモーダル音声行動コーパスの着想を得た．発話単位で，意図（ゴール）に基づく行動を，個別に表現するのが基本方法である．行動の外面的特徴のマルチモーダル観測に基づき，感情・意図・思考の内面的特徴の表現方法を追究する．

連絡先: 432-8011 静岡県浜松市中区城北 3-5-1 静岡大学情報学部

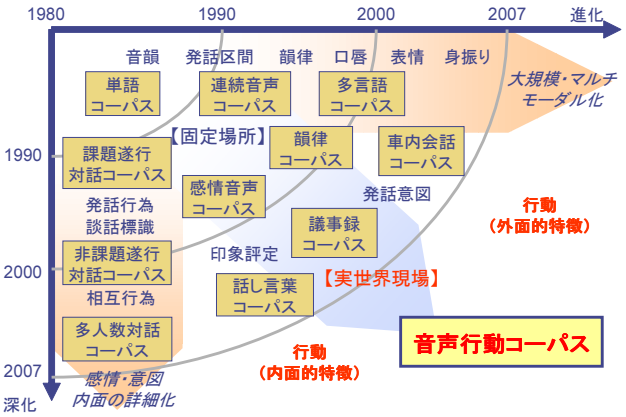

図 1: 音声コーパスの発展

思考の記述には，Minsky の提唱する多層思考モデル（図 2）を用いる．この図は，幼児期は下位の本能，資質，衝動，意欲により行動や思考が左右され，成長するにつれて自我が生まれ，成人になると，上位の価値観，理想，検閲，タブーという思想や倫理観が複雑に絡み合って，思考や行動に影響を与えることを示している．中間の層は，あらゆる種類の問題，衝突，目標を扱うための思考方法であり，日常生活における厖大なコモンセンス知識を含む [Minsky 06]．

3. 主観の客観化

感情・意図・思考の内面的特徴は解釈が多様で，主観によって仮説を記述するしかない．従来のコーパスは記述の正確さを追究してきたが，本コーパスでは，主観的な観察と洞察に基づく仮説を多面的な観点に照らして客観化するプロセスを重視し，記述の正当性を保障しない．一意に決まらない曖昧さを許容するからこそ，人間の多層的な思考の本質を考える材料になるのである．

図 2: 多層思考モデル

医者が患者の所見や治療方針を議論するカンファレンスを，本研究のコアメソッドとして用いる．類似の場面を複数取り出して考察，過去の同じ場面を時系列に並べて比較という分析を複数人で実施し，異分野の専門家，指導者，ユーザなど多視点からの見解を統合して最適な解釈を決定するプロセスで仮説を検証し，客観化する．

4. 幼児の音声行動コーパス

幼児はナイーブであり，大人に比べ感情・意図をストレートに行動に表出することが多いので，行動観測に基づく思考過程の考察に向いている．2005 年 6 月から静岡大学のキャンパス内で幼児教室を毎週開催し，保護者の賛同を得てコーパスを経年的に蓄積できる体制を整え，幼児の自発的な行動を映像と音声で収録できる世界に例を見ないマルチモーダル幼児教室環境を実現した．2004 年 10 月〜2008 年 3 月までに 234 回・326 時間分の映像・音声データを蓄積しており，同一幼児の経年変化の追跡が可能である．

幼児が単に楽しく遊ぶ場面などは，思考の考察に向かない．欲しいものが手に入らない，取り組みがうまくできない，など内面の考察に意味のあるアクシデントの場面に着目して，コーパスを構築した．これまでに，1 年分の幼児教室を詳細に観察し，発話単位で，発話内容，話者，発話相手，身振り，視線，表出された感情などの項目ごとに行動記述を蓄積してきた．まごとの役の取り合いで，女の子を泣かせてしまった男の子が，自分のやりたい役を諦めて女の子に譲るシーン，先生や母親の注意を引くために，韻律や身振り，視線を変化させて工夫するシーンなど，積み木のおもちゃを組み立てる際に，途中まで組み立てた積み木を壊して行き詰まりを打破し，先生のお手本を真似して先に進むシーンなど，本能，経験，社会性に関する多層的な思考に基づいた行動の事例を豊富に抽出できた．

5. 多層思考モデルの構築と検証

多層思考モデルは，人は常に複数の目標（ゴール）を持っており，ゴールの達成方法（問題解決方法）の候補から最適な方法を選択することにより行動を起こすという思想に基づき，ゴールの種類を，本能的，経験的，社会的など多層的に表現したものである．このモデルに基づいて，社会的思考による問題解決プロセスを表現・検証するために構築された EM-ONE 常識推論システム [Push 05] のソースコードを，MIT から入手している．MIT が持たない実世界データに基づく思考モデルを産出し，推論エンジンの改良に寄与する．EM-ONE が使用

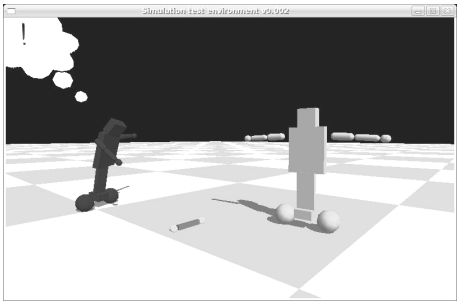

図 3: EM-ONE に基づく CG シミュレーション

する LISP ベースの知識表現言語を基盤として，感情・意図・思考の記述モデルを構築する．EM-ONE 知識表現言語は，信念，意図，ゴールの心的状況と，critic と呼ばれる問題解決知識を Actor ごとに記述するのが基本スタイルである．感情は，Actor が使用する問題解決知識を切り替えるスイッチである．感情音声研究の知見を踏まえ，感情と思考の関係を表現できる記述モデルを検討する．記述モデルの妥当性は，行動記述データを基に幼児の思考モデルを構築し，EM-ONE によるシミュレーション（図 3）で評価できる．

6. まとめ

マルチモーダル音声行動コーパスという新しい概念を提唱した．ナイーブな幼児の行動を対象にコーパスのプロトタイプを構築，分析して，人間の内面に踏み込んだ行動解析のための事例の宝庫であることを確認した．Minsky の多層思考モデルに基づく行動と思考の表現方法を検討し，CG シミュレーションによって検証という常識推論モデル構築の方法論を示した．様々な分野の研究者と交流しながら主観の客観化を推し進め，感情・意図・思考の表現モデルの進化を牽引する基盤を創出したい．

参考文献

[Furui 05] S. Furui, et al.: Analysis and recognition of spontaneous speech using Corpus of Spontaneous Japanese, Speech Communication, Vol.47, No.1-2, pp.208–219, (2005).

[Kawahara 08] T. Kawahara, et al.: Automatic lecture transcription by exploiting presentation slide information for language model adaptation, IEEE-ICASSP, SLP-L1.5, (2008).

[Irie 04] Irie, Y., et al: Speech intention understanding based on decision tree learning, INTERSPEECH-2004, pp.2185-2188, (2004).

[石黒 05] 石黒浩, ”アンドロイドサイエンス,”システム／制御／情報，Vol.49，No.2，pp.47-52, 2005.

[Minsky 06] Marvin Minsky, ”The Emotion Machine,” SIMON & SCHUSTER, (2006).

[Push 05] Push Singh, EM-ONE: An Architecture for Reflective Commonsense Thinking, PhD thesis, MIT (2005).