

マルチモーダル幼児行動コーパスからのコモンセンス知識抽出

竹林洋一[†] 桐山伸也[†] 大谷尚史[†] 山本剛[†] 川口紗季[‡]
仲川淳[†] Heikki Ruuska[†] 福島尚典[†] 北澤茂良[†]
[†]静岡大学 [‡]デジタルセンセーション株式会社

あらまし 本稿では、マルチモーダル幼児行動コーパスの構築を機軸とした幼児の思考の発達に関するプロジェクトについて述べる。行動コーパスは、幼児教室における幼児の行動を長期間定期的に収録した音声・映像情報と注釈データからなる。幼児教室の講師・研究者・保護者はコーパスを徹底的に活用して思考モデルの構築・評価・改良を SECI モデルにより協調して進め、幼児の活動や発話は見かけよりも複雑であり、Minsky の Emotion Machine 理論と行動コーパス構築がコモンセンス知識の研究に有用であることが示唆された。

キーワード マルチモーダル行動コーパス, 幼児教室, コモンセンス, 思考モデル

Commonsense Knowledge Extraction from Multimodal Infant Behavior Corpus

Yoichi Takebayashi[†] Shinya Kiriyama[†] Naofumi Otani[†] Goh Yamamoto[†] Saki Kawaguchi[‡]
Jun Nakagawa[†] Heikki Ruuska[†] Hisanori Fukushima[†] Shigeyoshi Kitazawa[†]
[†]Shizuoka University [‡]Digital Sensation Co., Ltd.

Abstract This paper describes an infant thinking development research project based on multimodal infant behavior corpus. Long-term SECI-model based collaborative analysis using the behavior corpus suggests that complicated infant commonsense thinking models can be created by multi-level commonsense reasoning theories in Minsky's Emotion Machine.

Key words multimodal behavior corpus, infant, commonsense, thinking model

1. はじめに

音声対話システムやロボットなど気の利いたインタラクティブ・システムの実現への期待が高まっており、かなりの規模の研究開発費を投じて、様々なインタラクション技術や多数のデモシステムが開発されるようになってきた[1-2]。しかしながら現状のインタラクティブ・システムのロバスト性は不十分であり、制約条件が強く応用範囲も限定されているため、本格的実用化の段階には至っていない。

ユビキタス情報環境下における人一人、人一人ロボットのコミュニケーションモデルの研究が盛んであるが、これらの研究は表

面的なマルチモーダル情報の解析評価や表層的なインタラクションモデルの検討レベルにとどまっており、人間の日常会話で必要な個々の人間の多面的で根源的なコモンセンス知識や意図感情モデルにまで踏み込んだ研究が必要とされている。

これまでの人間同士の音声の会話や人とロボットとのインタラクションの分析やデータベース構築の研究は、成人の言動や行動が研究対象であり[3-6]、幼児を対象にした長期的なマルチモーダル行動のコーパス構築やコモンセンス獲得の研究はなされていない。

以下本文では、筆者らが構築中のマルチ

モーダル幼児行動コーパスと、コモンセンス知識モデル構築へ向けた取り組みについて述べる。

2. 幼児の思考とコモンセンス知識

2.1 Emotion Machine としての幼児の思考

私たち人間は、目に入るものを捉え、耳に入る言葉を理解し、自然に会話をし、毎日の出来事や情報を記憶し、過去の経験や知識から様々な問題を容易に解決することができる。Minsky によると、私たちは著名な科学者や芸術家の能力を高く評価し尊敬するが、普通の人が日常生活で膨大なコモンセンス知識を駆使し状況に応じて行動し問題解決をしていることに気づいていない[7]。様々な環境下で人間のように柔軟で臨機応変に対応できる気の利いた音声対話システムや福祉ロボットを実現するためには、この人間の複雑で膨大なコモンセンス知識や卓越した思考モデルを構築することが必須であると考え、筆者らは幼児のコモンセンス知識の研究を進めることとした。

2.2 コモンセンス知識解明へのアプローチ

幼児の発話や行動は成人に比べてナイーブであり、複雑な思考過程が比較的外界にストレートに表出しやすい傾向がある。このような観点から、幼児の行動を多面的に観察することにより、幼児の思考及びその

発達過程を部分的にモデル化可能であると考え、幼児の心の健全な発達を目指す幼児の学習環境構築と並行して、幼児の根源的コモンセンス知識の研究を立ち上げた。

幼児教育の専門家の監修のもとで、大学構内で定期的に幼児教室を開きながら、幼児の発話や仕草の成長過程を観察し、幼児のコモンセンス知識の表現と獲得に関する基礎研究を行っている。

ヒューマノイド・ロボットの開発や人間とロボットとのインタラクションの研究、感情に関する研究が盛んであるが[8]、人間らしく振舞っていても、実際は人間の思考や感情とは全く違うメカニズムで動作しており、人間とロボットの共生には限界が来ると考えたからである。人間とロボットとのインタラクションも表層的な研究に留まっており、人間の思考モデル構築の検討は十分になされていない。音声対話システム構築のための基礎研究という視点ではコモンセンス知識と、それをを用いた常識推論の本格研究を推進すべきであると考えている。

2.3 幼児教育現場での知識創造

図 1 に筆者らの幼児の行動コーパスを機軸にした根源的コモンセンス知識の研究体制を示す。図 1 のサイクルは、野中らが提唱した SECI モデルに基づく知識創造[9]に対応している。幼児の行動を収集し、研究

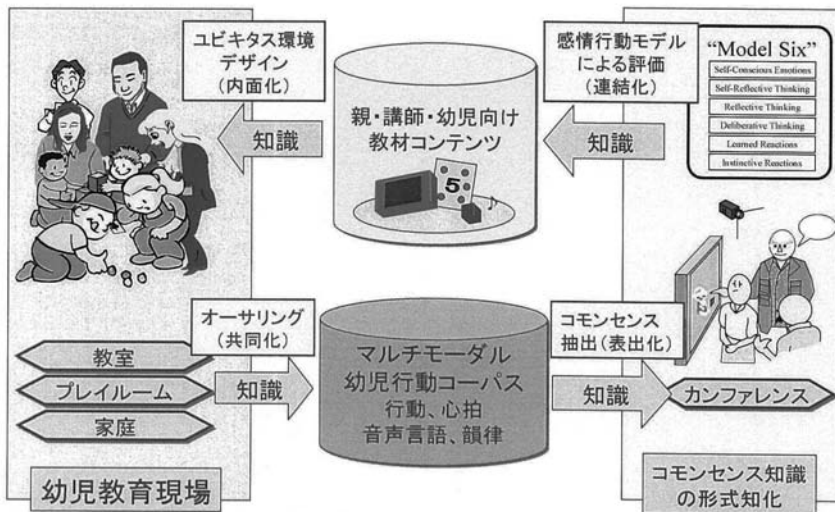


図 1: 幼児教育プロジェクトと知識創造サイクル

者が幼児の振る舞いや発話にアノテーションをつけて行動コーパスとして蓄積（共同化）する。研究チームで定期的にカンファレンスを開催し、音声言語やジェスチャ、他者とのコミュニケーションスキルの発達などの複数の観点からコーパスを分析し、コモンセンス知識を抽出（表出化）する。獲得されたコモンセンス知識は、幼児教育環境や教材コンテンツの開発・改良（連結化）に活用する。教材コンテンツは改善されながら幼児教室で使用され、幼児・親・教師の間の新たなインタラクションを創出（内面化）する。

図2はMinskyが提唱したコモンセンス知識による基本的なCritic-Selector（批評家-選択家）の思考モデルを示している。人間の日常行動では様々な場面で問題解決が必要となるが、解は必ずしも一つではなく正解や最善の選択肢が不明な場合が多い。人は問題に直面すると複数の仮説を生成して批判や評価を行い、試行錯誤を繰り返しながらあるアクションを選択する。経験豊富な人は、直面する問題を分析し、深く考えなくてもコモンセンス知識（常識）を使って判断し、最善の解決法を選択できる。その一方で経験の浅い人は、頭の中の多数の批判家の様々な意見を聞き過ぎて、収集がつかなくなり何も決められなくなり、試行錯誤を繰り返して判断が遅くなる。コモンセンス知識や経験が豊富かどうかで人間の思考や行動は異なるわけである。

3. 幼児教室の環境デザイン

3.1 幼児教室運営

筆者らは幼児のコモンセンスの知識の基礎研究と並行して、幼児の健やかな発達を実現するための実用的な研究を進めている。このため従来の認知科学や心理学のような特定の環境下で特定の刺激に対する幼児の反応を実験的に調べるといった研究の枠組み

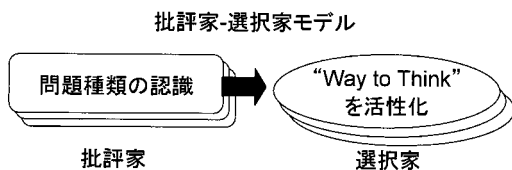


図2: MinskyのCritic-Selector思考モデル

ではなく、幼児の長期間にわたる成長や学習の観察を基盤に研究を進めている。幼児の成長に伴って新たな行動が表出してくる様子を捉えるため、図3に示すように学内に幼児教室を設置した。

学内に幼児教室を設置することで、幼児に与える教材、授業時間、教室内の空間の使い方を自由に決めることができ、幼児の発達過程が表出する場面を効率良く収集することを実現した。教室内で幼児の言動を記録するために、ネットワーク経由で遠隔操作可能なカメラ（4つ）と、固定の集音マイク（2つ）、個々の幼児の発話音声を収集するために幼児が常時着用可能な音声収録装置を配置した。

3.2 マルチモーダル幼児行動収録環境

この幼児教室環境を用いて、2005年6月から毎週1回幼児教室を定期的に開催してきた。1～3歳児クラスがあり（2007年3月現在）、それぞれ親子3組ずつが参加している。21ヶ月に渡って通算146回の幼児教室を開催し、217時間分の授業記録を映像・音声データとして収集してきた。

幼児教室中に親との育児相談時間を設けることで両親とも協力的な関係を構築してきており、幼児教室での授業時間外での出来事や幼児の変化も継続的に収集できる体制を実現している。収録データは子育て番組の素材としても活用している[10]。

4. マルチモーダル幼児行動コーパス

幼児教室の運営を通して収集した映像や音声データ、テキストデータによる幼児の行動記録を効率よく分析するため、以下の



図3: 幼児教室と行動観察環境

2 点を実現する幼児行動分析ツールを開発した。実行時のスクリーンショットを図 4 に示す。

- (1) 立場の異なる人が幼児行動記録にアクセスでき、多視点から幼児の行動分析に参加できること
- (2) 分析した結果を検証するために、分析結果を後から計算機処理可能な形で残しておくこと

4.1 行動記録ツールによる幼児行動観察

1 つ目の要求を実現するため、指定した複数の映像や音声からなる観察データをファイルサーバから読み出し、映像・音声データ中の望む箇所から再生できる機能、見ている映像と同期して再生する音声ソースを環境マイク・ウェアラブルマイクと自由に切り替えられる機能を実装した。

環境マイクの音声をういて幼児と他者とのやり取りを分析すること、および各幼児の発話能力の分析を同時に行える観察環境を実現した。

記録に対するコメント入力は、自然言語文で入力を可能にした。これによって誰でも記録データを再生しながら幼児の行動分析に参加できる。

4.2 コモンセンス知識の記述

Minsky の提唱する 6 階層思考モデル[7]に基づいてコモンセンス知識を記述する。図 5 に示すように、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い、思考や行動に影響を与える。こ



図 4: 幼児行動分析支援ツール

の思想に基づき、先天的・反射的な本能的思考、経験から学んだ学習的思考、相手への影響や他人の目を意識した社会的思考という複数の階層からなるコミュニケーションスキル発達記述モデルを設計し、知識記述を進めている。

Singh と Minsky が構築した EM-ONE アーキテクチャ[11]に基づいて幼児の行動や思考過程を表現・検証することとした。EM-ONE アーキテクチャは、(1)行動や思考を記述するための Lisp 形式に則った EM-ONE 記述言語、(2)EM-ONE 記述言語で記述した思考方法をもとに推論するプログラム、(3)推論結果をもとに CG 世界におけるロボットの行動をシミュレートするシステムから成る。

EM-ONE アーキテクチャのソースコードを基にして、実世界での様々な観点での観察事例と整合するように EM-ONE アーキテクチャを拡張し、現実の幼児行動の観察から得られた知見をモデルの高度化にフィードバックする。

EM-ONE の知識表現言語では「green (ロボット) は pink が板を持ち上げてくれることを望んでいる」というシナリオは下記のように表現できる。

```
(defnarrative partner-does-not-know-desire
  (together
    (desires green (does pink (lifts pink board))) [1])
    (desires green
      (believes pink (desires green (does pink (lifts pink board)))) [2]))
    (implies [1] [2]))
```

自然言語アノテーション・ツール上に、EM-ONE 記述言語に基づく記述作業を支援する機能を実装した。(図 4 下部)

自然言語文での記述に対して、対応する



図 5: Minsky の 6 階層思考モデル

EM-ONE 記述言語表記に書き直す作業を記述事例や記述のヒントを参照しながら行うことができる。

EM-ONE 記述言語の語彙では表記できない幼児の行動に対しては自然言語の表記で状況を記述して残すことができ、後から別観点での考察に利用できる。自然言語記述だけでなく、記述モデルの構築など行動記述の仕組みを高度化することにより、それに則った記述が多数で行うことができるようになり、行動分析を効率よく進められることが判明した。

5. カンファレンスによる記録の分析

5.1. カンファレンスによる議論

カンファレンスは医療分野の診断や治療方針の決定など複雑な問題解決の場で重要な役割を担っている。コンセンサス知識や思考モデルのデザインは様々な観点からの議論が不可欠であり、プロジェクト関係者による週 1 回の定期カンファレンスの場を設けた。発話やジェスチャ、視線、表情などの情報を手がかりに、各人が独自の観点で観察結果に基づく仮説を持ち寄り、議論を通して仮説の検証を進めている。

前述した幼児行動分析ツールと行動コーパスを実装後、収集した 3 歳児クラスの 2006 年度の授業 1 年分、20 時間分の授業データを対象として 600 の場面に自然言語文によるコメントを付与した。このコメントデータを用いて、収集したコメントデータの分析をカンファレンス形式で分析を行った。ネットワークを介して参加者全員で記録データを共有し、参加者全員で授業中の幼児の行動記録を見ながら 2 ヶ月間に亘って毎日 1 時間 5 人がカンファレンスを行った。

幼児がゴール達成に向けて問題解決に取り組む様子の中から、幼児を 1 ヶ月毎に観



図 6: 幼児の試行錯誤の事例
(相手の注意を引くことができなかった場合指差しではなく直接触れる手段に変える)

察し、意図に応じて表出する発話内容、目線、身振りが記録されていた 240 のシーンについて議論した。

5.2. 幼児の行動モデルの検証

幼児行動コーパスから着目する場面を抜き出し、例えば図 6 に示すような場面に対してカンファレンスを行い、思考モデルについて議論を行った。相手の気を引くために月齢 16 ヶ月目で指差しをするだけだったのが、月齢を重ねるにつれて相手を叩くという行動に変化していた場面、視線が 16 ヶ月の時点では幼児自身の注目対象に固定だったものが、月齢を重ねるにつれて注意を引きたい相手を見るように変化していく場面をカンファレンスの場で議論した。この結果をもとに、指示表現における月齢別の行動モデルを構築した。

構築した幼児行動モデルを検証するため、EM-ONE システム中のシミュレーション環境上のロボットに発話機能を導入し、指差し・手差しを表現できるロボットシミュレーション環境 (roboverse II) を実装した。動作中のアニメーションのスクリーンショットを図 7 に示す。

実例をベースにモデルを構築し、シミュレーション環境で動作させ、観察することで構築したモデルの検証を行える見通しを得た。

6. おわりに

幼児教室を定期的で開催しながらマルチモーダル幼児行動コーパスを構築し、幼児

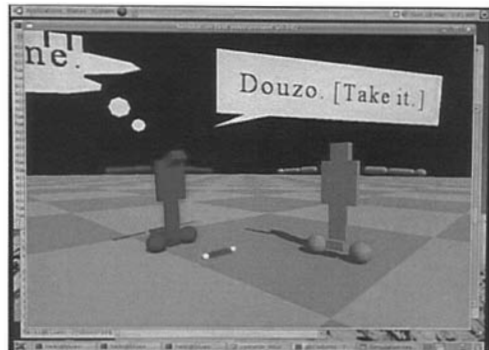


図 7: 幼児の多階層思考も出るに基づく行動シミュレーション環境

行動分析と多階層コモンセンス思考モデルを評価改良できる環境を開発した。気の利いた音声対話システムや気の利いたロボットを実現するためには、音声言語処理の研究に加えて人間の意図感情やコモンセンス知識の研究の高度化が必須である。

今後、幼児教室を開催しながら音声言語の発達や概念獲得、さらにはジェスチャ、表情、コミュニケーション能力などのマルチモーダルコモンセンスの獲得に関する研究を進めるとともに、2006年10月に設立した「幼児のコモンセンス知識研究会[12]」を主体に異分野の研究者と交流しながら根源的コモンセンスの基礎研究を深化発展させていきたい。

参考文献

- [1] 竹林洋一: 音声自由対話システム TOSBURGII ―ユーザ中心のマルチモーダルインタフェースの実現に向けて―, 電子情報通信学会論文誌, Vol.J77D-II, No.8, pp.1417-1428 (1994).
- [2] 竹林洋一:「ヒューマンインタフェースの進化」,第 50 回 音声言語情報処理研究会(SIG-SLP),ホテル池田(熱海) (2004.2.6).
- [3] 角康之 他: “協調的なインタラクションの記録と解釈,” 情処学論, Vol.44, No.11, pp. 2628-2637 (2003).
- [4] 坊農真弓 他: “ユビキタスセンサによるインタラクション・コーパスの収集とインタラクション分析への応用,” 情処大会, T7 (2006).
- [5] Roy, D., Patel, R., DeCamp, P., Kubat, R., Fleischman, M., Roy, B., Mavridis, N., Tellex, S., Salata, A., Guinness, J., Levit, M., and Gorniak, P.: *The Human Speechome Project, the Proceedings of the Twenty-eighth Annual Meeting of the Cognitive Science Society* (2006)
- [6] 西田佳史, 本村陽一, 山中龍宏: 乳幼児事故予防のための日常行動モデリング, 情報処理, Vol. 46, No. 12, pp.1373-1381 (2005)
- [7] Marvin Minsky : *The Emotion Machine*, Simon & Schuster (2006)
- [8] 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)
- [9] 野中 郁次郎, 竹内 弘高, 梅本 勝訳: 知識創造企業, 東洋経済新聞社 (1996)
- [10] Digital Sensation 株式会社: ぐんぐん子育て番組ホームページ, <http://www.digital-sensation.jp/dslive/mirura/>
- [11] Singh, P.: *EM-ONE: An architecture for reflective Commonsense Thinking*, PhD Thesis, Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology (2005)
- [12] 幼児のコモンセンス知識研究会:
<http://minny.cs.inf.shizuoka.ac.jp/SIG-ICK>