

## Open Challenges in Modelling, Analysis and Synthesis of Human Behavior in Human-Human and Human-Machine Interaction

*Cogn Comput (2015) 7, 397-413*

Alessandro Vinciarelli, Anna Esposito, Elisabeth Andre, Francesca Bonin, Mohamed Chetouani, Jeffrey F. Cohn, Marco Cristani, Ferdinand Fuhrmann, Elmer Gilmartin, Zakia Hammal, Dirk Heylen, Rene Kaiser, Maria Koutsombogera, Alexandros Potamianos, Steve Renals, Giuseppe Riccardi & Albert Ali Salah

### 1. イントロ

- コンピューティングと心理学、認知科学を合わせて検討
    - ✓ 人間の行動のモデリングや分析、合成はコンピューティング・サイエンスにおける主題の一つである
    - ✓ 人間の行動は、観察可能ではないが、認知的、社会的、心理的な現象によって支配されている。
    - ✓ そのため、より効果的な人間中心の技術を構築する必要がある
    - ✓ その例として、1990年代初頭に行われた顔の表情の自動分析に関する研究では観察可能な顔の筋肉の活動だけでなく、感情表現の観点からも初めて取り組まれた。
- 学際的に、こういったコンピューティングと心理学、認知科学を合わせて検討することは重要かつ実りあるものとされる。

### 2. データ

- コーパス作成とデータ収集は人間の行動のモデリング、分析、合成に必要な前提条件
  - ✓ 実際、分析と合成はデータからの学習なくして実現しない
  - ✓ 実験の効率化と結果の再現性は充実したデータアノテーションと分析のためのフレームワークがなければ困難

## データのデザインと収集

- マルチモーダルなインタラクション行動を示すデータの収集は言語的、非言語的、社会的なコミュニケーション現象を広くカバーしなくてはならない。
- しかし、現実のリソースを見てみると、社会的インタラクションのすべての側面に対処するのではなく、特定の文脈や設定の調査に焦点を当てている。
  - ✓ その理由は、口述インタラクションの範囲の莫大さにある。
  - ただ、話者交替やバックチャネル、ジェスチャーなどが多様化するのオープンクエスションの時であることを忘れてはいけない。

## コーパス作成では実験室で自然な対話を

- 実験室では自然な対話ができているか疑問
  - ✓ 実験室でインタラクションが記録される場合、タスクを使って対話させることが一般的
    - 例) 地図上のルートの記述 (HCRC Map Task Corpus)、類似した画像間の違いの発見 (London UCL Clear Speech in Interaction Database)、実際のまたはシミュレートされた専門家会議への参加 (ICSI Meeting Corpus)
  - ✓ この方法で台本的でない対話の有用なコーパスが得られる。
  - ✓ しかし、人工的なタスクに参加している参加者の動機は真なものであるか疑問
    - 得られたコーパスが自然な対話の一般化を議論するのに利用できるか不明。
- 自然な状況の対話からコーパス作成は難しい
  - ✓ より自然な環境でデータを収集しようとする試みとしては、電話を媒介とした会話を記録しようすることが一般的である (記録しやすい)。
  - ✓ しかし、これはユニモーダルで、これと対面での会話が同等であるかは不明。
  - ✓ さらに自然な対面での対話からコーパスを作成するのは比較的低品質の録音
    - コーパスとしてはモダリティや品質の面から制限されてしまう。
- 実験室で自然な対話を
  - ✓ 以上のことを踏まえると、コーパス作成に当たっては実験室実験が選択される
  - ✓ ただし、記録装置はなるべく目立たないように
  - ✓ シナリオはバイアスがかからないように
    - 参加者に課せられるタスクや議論の対象は決まってないほうが望ましい

### 3. アノテーション

- アノテーションは大変
  - ✓ 記述的かつ意味的な情報でデータ（コーパス）を扱うためには通常手動でのアノテーションが必要
  - ✓ 確かに近年、自動化も進んでる。
  - ✓ しかし音声、ジェスチャー、その他のマルチモーダルインタラクションの特徴を取り入れた包括的なアノテーション方法は定まってないのに自動化とは？  
→結局、研究に合わせて手動でアノテーション

### 4. 行動分析

- 個人の行動から対人への影響の検知・理解を図る
  - ✓ 従来の社会行動分析の研究では、個人に焦点を当てていた。  
※特定の行動やキュー（例：表情、ジェスチャー、プロソディー）の検出  
心理的現象（例：覚醒、性格特性）の測定
  - ✓ 近年、個人の行動検出を超えて、対人への影響の検出とその理解に関心高まる

#### 対人影響のモデリング

- 対人への影響を分析する2つのメインアプローチ
  - ✓ 分析的/記述的モデル：個人の行動が他の人の行動をどの程度説明するかを定量化
  - ✓ 予測/分類モデル：対話パートナー間の行動マッチングを測定
- 分析的/記述的モデル（パートナー間の相互関係を分析）
  - ✓ インタラクションにおける2人の「同調」を見る。  
→窓付き相互相関（Windowed cross-correlation）は、2つの時系列間の類似性を見る尺度として最もよく使われる。
  - ✓ 別のアプローチとして、リアルタイム実験パラダイム  
→ある人物の行動のアウトプットをアクティブアピランスモデルを用いて処理し、リアルタイムでフィードバック。どんな行動が返るか測定。  
例) アバターの頭のうなずきを減衰させると、相手の頭のうなずきと横向きの頭の回転が増加する

- 予測/分類モデル
  - ✓ 近年の研究では、パートナー間の類似した行動の瞬間を検出（予測）することに関心
    - 例) バックチャネルのような相互または、同期的な頭のうなずきを学習し、その発生を自動で検出しようとする。
  - ✓ この発生の瞬間を検出する方法として、半教師あり学習を用いた方法が提唱
  - ✓ 同様に、教師ありまたは教師なしの方法を用いて、パートナー間の話者交替のタイミングを検出する方法も示されている
    - よりよいエージェントづくりに役立つ

## 5. 会話エージェントや社会的ロボットとの人間のインタラクション

- 社会的シグナルは文脈に強く依存
  - ✓ 例えば、死別の状況で行われるジェスチャーは、同情を示していると簡単に解釈できるが別の文脈では全く不適切な場合もある。
  - ✓ 人間はこの認識と判断に長けているが、人工的なエージェントは微妙なニュアンスに対応するための意味的背景知識が不足している
    - この欠点を補うために、エージェントは（社会的文脈と独立した）明確で単純な信号で内部状態を人間に伝える必要がある。その逆も然り。
- 人-人インタラクションで使用される信号は少ない
  - ✓ 例えば、ペットの猫は人間とは異なる社会的なシグナルを生成するが、この一連のシグナルを介して人間とスムーズにコミュニケーションを行う。
  - ✓ Audio-Visual Emotion Challenges (AVEC)のような国際的なベンチマークキャンペーンで、人間とエージェントのインタラクションでは、必ずしも人間と人間のインタラクションで使用される信号と同じ信号を使用する必要がないことが示された

※もちろん文脈の中での認識、交渉、社会的手がかりの生成のための努力は必要

## 6. インタラクションのコンピューテーショナルモデル

- 人間のインタラクションの3段階プロセス (Tomasello)
    - ✓ 注意の共有
    - ✓ 共通基盤の確立
    - ✓ 共有目標の形成 (共同意図)
- 認知機能の面から
- ✓ 低: (共同) 注意を介したキューの選択
  - ✓ 中: 共通基盤を介した意味の曖昧性解消
  - ✓ 高: 意図認識

### 注意

- キュー選択のための注意
  - ✓ コンピュータとは異なり、人間は画像や音、パンフレットなどの最も顕著な部分のみを処理。
  - ✓ この注意を利用して、パートナーからのキューを選択
    - 人間が何を見聞きするかをモデル化して予測することができるようになることは、エージェントが認知的表現を形成するための第一歩
- 注意モデルはトップダウン・ボトムアップ両側面が必要
  - ✓ しかしながらこれまで提唱されてきたインタラクション注意モデルは、知覚的な手がかりを利用した (ボトムアップ) 低レベルなもので高レベルな意味情報を利用 (トップダウンの検討) していない。
  - インタラクションの注意モデルを構築することはこの分野の未解決問題

### 共通基盤

- 人間は相互作用しながら、低・中・高レベルの認知機能を統合
    - ✓ 低: (共同) 注意を介したキューの選択
    - ✓ 中: 共通基盤を介した意味の曖昧性解消
    - ✓ 高: 意図認識
- 共通基盤を確立させることがコミュニケーション成功の前提
- 理想的な対話マシンは、パートナーとの共通基盤となる概念とその関係のネットワークをモデル化すべき
    - ✓ フォーマルオントロジーが良いのでは。
      - ◇ トップダウンの意味表現
      - ◇ 数学的論理を介した推論力

## 7. 結び

- ① 人間の行動は常に状況に応じたものであり、文脈に依存しているということである。
  - ✓ 人間の行動に対処するための技術は汎用的なものにするのではなく、使用される文脈の非常に特異的な側面に対処しようとするべき。
- ② 人-人と人-機械のインタラクションの言語的側面と非言語的側面の両方を考慮する必要性
  - ✓ これまで言葉の内容と意味論は軽視されがちだった
  - ✓ その理由は、非言語的な側面の方がより正直であり、さらに、人々の発言を考慮に入れることはプライバシーを侵害するから
- ③ 人間の対話的行動を駆動するプロセスを明示的にモデル化すること