

# 評定者個人に特化した他者感情理解モデル

## Perceiver-specific Affect-Understanding Model

熊野 史朗\*<sup>1</sup>   石井 亮\*<sup>1</sup>   大塚 和弘\*<sup>1</sup>  
Shiro Kumano   Ryo Ishii   Kazuhiro Otsuka

\*<sup>1</sup>日本電信電話株式会社

Nippon Telegraph and Telephone Corporation

本稿では、ある個人が他者の感情状態をどう理解・評定するかを予測するための数理モデルを提案する。特に、評定者の特性、及び、被評定者が表出する観測可能な非言語行動という二要因に着目し、両要因の独立性を仮定してそれぞれ独立にモデル化する。評定者モデルは、ある評定者の評定傾向を説明するモデルであり、ここに被評定者の情報は含まれない。本稿では評定者の性格特性と性別を入力として利用する。他方、対話行動モデルは、ある行動が不特定の評定者からどう評定されそうかを説明するモデルである。本稿では表情や視線等の非言語情報の共起情報とタイミング情報を入力として利用する。それぞれのモデルを各評定値の確率を出力する統計モデルとし、それらをベイズ則に従い統合することで、ある評定者と評定対象の対話行動（シーン）が与えられたもとで認知される感情の条件付き確率を計算し、最大確率の評定値を推定結果とする。様々な対話シーンに対して多数の評定者が5段階で与えた独自の共感認知データセットに対して、再テスト信頼性を1として7割程度という高い推定精度を得た。

## 1. はじめに

複数人対話の機械支援やエージェント構築に重要となる対話者同士が織り成す感情の場を自動で理解するためには、各人物自身の内部に実際に生じている感情状態と、それが各々の他者からどう認知されるかの両方を理解する必要がある [1]。両者はしばしば異なり、誤解は望ましくない結果に結びつきやすいためである。だが、これまでの研究の大半は、前者の構築を対象とする、あるいは、後者に関して、不特定の人物、言い換えると集団にどう認知されるかを対象としており、各個人が他者の感情をどう認知するかを扱う研究は非常に少ない。

その主な理由の一つは、他者感情認知には、ばらつきを大きく含む多数の説明変数が存在し、かつ、それらの交互作用が存在するためと予想される。例えば、認知する人物の個人差、感情理解の手がかりとなる認知される人物の状態（言語・非言語情報など）、両者の関係、観測の環境（例えば、実際に対話の場の中にいるか、外から一方的に眺めているだけなのか）などが挙げられる [2]。よって、これら全ての要因を考慮した精緻なモデルの構築には膨大なデータが必要であり、Web上の様々なデータやクラウドソーシングを活用できる現在においても依然として容易ではない。

そのため、本研究では、評定者の特性、及び、被評定者のとる観測可能な対話行動という二要因を扱う。さらに、それらの交互作用を無視する。これにより、評定者と被評定者の対話行動を独立にモデル化することができ、小規模なデータからでも過学習を抑えたモデル獲得が可能となる。この仮説の妥当性については3.4にて検証する。評定者のモデルは、ある個人が、ある行動をどう評定しそうかを表現するモデルであり、評定者のスキルやバイアスを含んだ認知傾向を表す。対話行動モデルは、被評定者のある対話行動が、評定者「群」、すなわち、未知のある評定者に対して、どのように評定されるかを表現するモデルである。

本稿では、我々独自の共感認知データコーパスを用いて、提案モデルの有効性を確認する。だが、提案方法は汎用的であり、数理的には他の様々な種類のデータに対して適用可能である。

## 2. 提案モデル

評定者  $x_1$  が、対話行動  $x_2$  を表出する不特定の人物に対して与える感情の評定値  $y$  である。我々は評定値  $y$  の推定に確率推論の枠組みを用い、評定者  $x_1$ 、対話行動情報  $x_2$ 、及び、感情認知の評定値  $y$  の同時確率  $P(x_1, x_2, y)$  を最大化する  $y$  として推定する。ここで、評定値  $y$  が与えられたもとの評定者  $x_1$  と対話行動情報  $x_2$  の独立性を仮定する、すなわち、 $P(x_1, x_2, y) = P(x_1|y)P(x_2|y)$  とする。これは、工学的（例えばパターン認識）的にはナイーブベイズモデルを用いることと等価であり、過学習を防ぐ一つの方法として広く用いられている。他方、心理学的には両説明要因の交互作用が存在しないことを仮定していることに近い。このとき、途中式は省略するが、同時確率は最終的に下記の形に分解される。

$$P(x_1, x_2, y) \propto P(y|x_1)P(y|x_2)/P(y). \quad (1)$$

ここで、 $P(y|x_1)$  は評定者（の認知傾向）モデル、 $P(y|x_2)$  は対話行動モデル、 $P(y)$  は評定値の事前分布である。

本稿では評定者モデル  $P(y|x_1)$  として、評定者の属性情報（性格特性に関する心理尺度スコアと性別）と認知傾向とを対応付ける確率的トピックモデル [3] を用いる。また、対話行動モデル  $P(y|x_2)$  として、非言語行動情報、より具体的には、表情、視線、頭部ジェスチャ、発話の有無に関する二者間の共起パターンとそのタイミングを用いるモデル [4, 5] を用いる。両モデルにより、未知の評定者  $x_1$  や未知の行動場面  $x_2$  に対する推定が可能となる。

## 3. 実験

### 3.1 対話（刺激）データセット

我々が過去に撮影した5組の4人対話データをもとに、独自の対話刺激データセットを用意した。いずれの対話も、与えられた議題に対してグループで一つの結論を出す議論型の対話であり、各議論の長さは8分前後であった。我々は、まず、この各対話に対して、各時刻における各二者の間の共感状態を5-9名の女性により付与した [4]。共感状態の定義については、共感を二者の感情が類似した状態、反感は二者の感情が異

なる状態として、「共感」「やや共感」「どちらでもない」「やや反感」「やや反感」の5状態を用意した。さらに、対話行動情報として、表情(6カテゴリ)を3名の女性による多数決により、視線(被注視人物を表すカテゴリ)、頭部ジェスチャ(頷き、傾げ、首振り)、及び、発話(有無)を1名の女性により付与した[4, 5]。なお、これらいずれの人物も本稿における評定者とは異なり、また、いずれのアノテーションもフレーム単位で音声なしで行った。

次に、これら共感情報、及び、対話行動情報を用いて、2名が映る6秒間のシーン97つを、1)(視線行動が理解しやすい)両名が横並びに座っており、2)何らかの対話行動の変化が見られるシーンであり、かつ、3)5-9名が与えた共感ラベルの頻度分布がなるべく様々な形状をカバーするよう自動的に選択した。これらの97シーンの映像を、提示順序についてカウンターバランスをとったうえで、評定者に音声無しで二回ずつ提示した。さらに、97つのうち3シーンについては、再テスト信頼性の計算のため、それぞれ最後にもう一度提示した。なお、対話行動情報については、対話行動モデルの入力としても用いた。

### 3.2 評定者と評定方法

100名(女性50名)の20代の大学生が実験室にて参加した。全員が対話者と面識がなかった。全評定者は3.1と同じ共感状態の定義のもと、5肢強制選択で、全シーンに対して評定を一人で行った。さらに、各評定者は、評定者モデル構築用として、評定タスク終了後に心理尺度に関する質問に回答した。ここでは、Davis' Interpersonal Reactivity Index (IRI), Emotional Skills and Competence Questionnaire (ESCCQ), 及び、東京大学エゴグラム (TEG) の3つを用いた[3]。

### 3.3 評価方法

推定精度の評価尺度としては、実際の推定精度を再テスト信頼性で除して正規化した値を使用する。例えば、正規化された推定精度が1であれば、再テスト信頼性と同一精度であることを意味する。再テスト信頼性は、デブリーフィングを行い、反復に気がつかなかったとの回答が得られたサンプルのみから算出することとする。

また、推定方法については、評定者モデルと対話行動モデルの両方を用いた場合に加えて、それぞれ単体の効果を比較するため、 $P(y|x_1)$  と  $P(y|x_2)$  を、モデルからではなく、対象評定者が与えた/対象シーンに与えられた評定値から頻度分布として直接計算した分布を真の分布のよい近似解として、両モデルと取り替えたモデルも評価する。

ここではその近似解を一個抜き交差検定で得ることとする。すなわち、評定者モデルの場合、推定対象の評定者が推定対象のシーン以外の97-1つのシーンに対して与えた評定値の頻度(確率)分布である。対話行動モデルの場合、推定対象のシーンに対して推定対象の評定者以外の100-1名の評定者が与えた評定値の頻度(確率)分布である。これら近似解とモデルとの違いは、前者は対象評定者/対象シーンの評定値の情報を必要とする(それゆえ未知の評定者/シーンに直ちに適用できない)のに対して、後者はそれらを必要としない(それゆえ未知の評定者/シーンに適用可能な)点である。

### 3.4 結果

3シーンそれぞれについて反復に全く気がつかなかったとの回答が得られたサンプルは全体の三分の二であり、再テスト信頼性は0.53であった。5肢強制選択(チャンスレベル0.2)、及び、二者感情の類似性という単純でない課題から考えると決して低い値ではないと考える。

表 1: 推定精度

| 方法 | 評定者モデル     | 対話行動モデル | 精度   |
|----|------------|---------|------|
| 1  | 近似解        | 近似解     | 0.99 |
| 2  | モデル [4, 5] | 近似解     | 0.86 |
| 3  | 近似解        | モデル [3] | 0.82 |
| 4  | モデル [4, 5] | モデル [3] | 0.68 |

再テスト信頼性が1になるよう正規化している。

提案法の推定精度を表1に示す。まず、方法1、すなわち、評定者モデルと対話行動モデルを共に近似解で置き換えた方法が0.99という非常に高い精度を達成した。ここから、提案法の根幹である独立性の仮定(式1)は妥当であったと考える。次いで、評定者モデルと対話行動モデルを共に用いた場合、すなわち、方法4の推定精度は7割程度であった。評定者モデル、あるいは、対話行動モデルの片方のみを用いて、残りを近似解とした場合、すなわち、方法2及び3については、8.5割程度であった。先行研究の少ない評定者モデルのみを用いた場合の精度が、先行研究のより豊富な対話行動モデルのみを用いた場合の精度よりも若干高いのは、二名の対話者で構成される対話行動のばらつきの方が、一名の評定者のみで構成される評定者モデルのばらつきよりも大きかったためと考える。

## 4. まとめ

本稿では、ある個人が他者の感情状態をどう理解・評定するかを予測するための数理モデルを提案した。特に、評定者の特性、及び、被評定者が表出する観測可能な非言語行動という二要因に着目し、両要因の独立性を仮定して、ベイズ則に従いそれぞれ独立にモデル化する方法を提案した。我々独自の共感認知データセットを用いた評価では、このアプローチそのものの妥当性は強く示唆された。だが、対象評定者が与えた評定データや対象シーン(対話行動)に与えられた評定データを用いることなく、推定しようとした場合、推定精度の向上に大きな余地が残されている。これは、深層学習等のより高度な機械学習技術の導入による改善の余地が多く残されていることを示唆する。

## 参考文献

- [1] R. Cowie and R. R. Cornelius, "Describing the emotional states that are expressed in speech," *Speech Commun.*, vol. 40, no. 1-2, pp. 5-32, 2003.
- [2] D. A. Kenny and L. Albright, "Accuracy in interpersonal perception: A social relations analysis." *Psychol. Bull.*, vol. 102, no. 3, pp. 390-402, 1987.
- [3] S. Kumano, K. Otsuka, M. Matsuda, R. Ishii, and J. Yamato, "Using a probabilistic topic model to link observers' perception tendency to personality," in *Affect. Comp. Intell. Interaction (ACII)*, 2013, pp. 588-593.
- [4] S. Kumano, K. Otsuka, D. Mikami, M. Matsuda, and J. Yamato, "Analyzing interpersonal empathy via collective impressions," *IEEE Trans. Affect. Comp.*, vol. 6, no. 4, pp. 324-336, 2015.
- [5] S. Kumano, K. Otsuka, M. Matsuda, and J. Yamato, "Analyzing perceived empathy based on reaction time in behavioral mimicry," *IEICE Transactions*, vol. 97-D, no. 8, pp. 2008-2020, 2014.