

特集 「人工知能と Emotion」

対話システムと感情

Dialogue Systems and Emotion

東中 竜一郎
Ryuichiro Higashinaka
日本電信電話株式会社 NTT メディアインテリジェンス研究所
NTT Media Intelligence Laboratories, NTT Corporation.
higashinaka.ryuichiro@lab.ntt.co.jp

岡田 将吾
Shogo Okada
東京工業大学情報理工学院
School of Computing, Tokyo Institute of Technology.
okada@c.titech.ac.jp

藤江 真也
Shinya Fujie
千葉工業大学先進工学部
Faculty of Advanced Engineering, Chiba Institute of Technology.
shinya.fujie@p.chibakoudai.jp

森 大毅
Hiroki Mori
宇都宮大学大学院工学研究科
Graduate School of Engineering, Utsunomiya University.
hiroki-public@speech-lab.org

Keywords: dialogue systems, emotion, text processing, speech processing, multimodality, robotics.

1. はじめに

スマートフォン上で動作する対話システムが一般ユーザに利用されるようになってきた[中野 15]. これらのシステムは主に会話によって所定のタスクを遂行するタスク指向型対話システムであるが, 近年は日常会話, いわゆる, 雑談をユーザと行うシステムも増えてきた. しかしながら, 我々とシステムの会話が人間同士のそれと同程度に自然というまでにはまだまだ先が長いのが現状である. これは, 音声認識や言語処理の能力のみならず, その場の状況や, ユーザの状態・感情を理解する能力も十分でないからである. 特に, 対話システムにおける感情は工学的な取組みが難しくあまり研究が進んでいるとはいえない. しかし, 感情のない人間と話をしたくなるだろうか?

そもそも, 我々が望む感情をもつ対話システムとはどのようなものだろう. いつも感情をむき出しにするようなものだろうか? 考えてみると, 一般に感情は内面にあるもので, それをそのまま表出して会話する大人はいない. 人が感情をむき出しにしているのはせいぜい幼児期までで, 後述するように感情的な振舞いを扱うロボットが幼児や動物をモデル化したものが多いということからも, 感情をモデル化したとしても, それをそのまま出力するシステムは, ペットロボットのように愛玩や癒やしなどの, 限定されたタスク以外では求められてはいないと思われる. よって, 我々が理想としているものは, 「感情むき出しではないが, 感情をもった対話システム」と

いうものだと考えられる. そのようなものが何の役に立つのか. 我々が考えるところによれば, 利点は次の三点である.

一つ目は, ユーザの感情をくみ取ることができるということである. 「顔では笑っているが心では泣いている」ことや, 「言葉では『大丈夫』と言っているけれど, 顔・表情では助けを求めている」ことがわかるようになる.

二つ目は, システムの状態をより直感的に示すことができるということである. 現在の対話システムの問題点の一つは, システムが何を考えているかわからないことであるが, 例えば, タスク指向会話におけるタスクの困難さ(無茶な要求をされているかどうか)について, システムがどう捉えているかをいったん感情状態にマッピングしたうえで, シャベリ方や表情などを変化させることにより, 間接的にユーザにフィードバックすることができるようになる.

三つ目は, 機械との対話をより自然で豊かなものにできるということである. 感情は人間同士の社会的関係の形成に重要な役割を果たしている. これを対話システムが利用することにより, ユーザとのより良い社会的関係を構築できる可能性がある. 社会的関係の形成に重要な現象として, 感情の同調(emotional contagion)[Hatfield 93]があげられる. 音声対話システムにおける同調現象の利用については3.2節に記述する.

このような利点を考えれば, 感情はこれからの対話システム, 特に, 日常会話を行うような対話システムにとって重要な要素となるだろう.

本稿では, 対話システムと感情に関する研究をサーベ

的に紹介することで、対話システムの研究者や開発者が自身のシステムに感情の要素を取り込むきっかけを与えることを目的とする。まず、感情の定義と類型について触れ、その後、対話システムの主要分野である、テキスト処理、音声処理、マルチモーダル処理、ロボティクスの各分野において、感情がどのように扱われているかについて述べる。その後、対話システムにおける感情モデルについて説明する。

2. 感情の定義と類型

感情（あるいは、情動、emotion, affectとも呼ばれる）という語を明確に定義することは困難である [森 14]。問題を整理しづらくしている一因は、人々が「感情」という便利なスツケースワード [Minsky 07] に押し込んだ概念が、あまりにも多様かつ曖昧なことにある。したがって、対話システムやその要素技術において感情を扱っていると標榜している研究は多く存在していても、それらが実際に指すものはさまざまである。

心理学における感情の理論には、主に二つの考え方がある。一つはカテゴリカルな感情を前提とする基本感情説、もう一つは、連続的な感情を前提とする次元説である。それらをまず紹介しておく。

感情を進化の産物であるとする進化論学派は、基本感情と呼ばれる先天的かつ環境や文化に依存しない普遍的な感情の存在を仮定する。基本感情のカテゴリーは研究者によって異なり、Plutchik は八つの基本感情軸を仮定し、それらに属する感情語の強度を質問紙調査によって求めた。これは感情の輪として知られている。また、Ekman らは、怒り、嫌悪、驚き、幸福、恐れ、悲しみの 6 感情をあげている [Ekman 72]。音声や表情からの感情分類では、これらのいわゆる「6 大感情」をカテゴリーに含めることが多い。

基本感情説に対し、感情の次元説では基本感情の存在を認めない。感情のカテゴリーは感情空間における位置を示すだけであり、重要なのはその空間を定義する次元であると考えられる。Russell [Russell 80] は「快-不快」と「覚醒-睡眠」の二次元による円環モデルを提唱した。これらの二次元は、感情の基本的な次元として広く受け入れられている。また、第 3 の次元として「支配-服従」を加えることも多く、Plesantness-Arousal-Dominance (PAD) モデル [Mehrabian 74] と呼ばれることがある。

対話システムは上述したような感情の定義や類型に従って動作するものが多い。しかし、その反面、インタラクション場面で重要な役目を担う社会的シグナルは、怒りや悲しみといった典型的な感情ばかりではない。感情の定義を狭く捉え過ぎることは、対話システムの高度化の観点からは得策とはいえないだろう。安らぎ、安堵感、信頼、親しさ、愛着、ストレスなどの周辺的な事象のほうが、対話システムとしてはむしろ重要なこともある。

3. 各分野における感情の扱い

3.1 テキスト処理

テキスト処理における感情は、書き手の感情、読者がテキストを読むことで受ける感情、テキスト自体に含まれる感情の 3 種類があるとされ、それぞれの観点から感情を認識する技術が取り組まれてきた。なお、感情と言っても、いわゆる基本感情を認識するという研究は少ない。なぜなら、テキストでは怒りや驚きといった感情はどうしても伝えることが難しく、テキストに現れにくいからである。このことから、テキスト処理における感情は、極性（ポジティブ・ネガティブ）、もしくは、これに類する軸で展開されてきた。

テキスト中に含まれる感情については最も多く取り組まれてきている。商品レビューや Twitter のような consumer generated media (CGM) を対象として、その極性を推定するものである。これらの研究は、sentiment analysis、もしくは、評判分析と呼ばれ、マーケティングへの有用性から、言語処理の一大分野となっている。初期の研究は、テキストの極性を推定するものが主だったが [Turney 02]、その後、レビューの星の数を当てるような極性の度合いを求めるものや [Pang 05]、商品の属性とその評価値をセットで求める研究も多くなされている [Popescu 07]。また、文章全体に極性を付与するのではなく、文章中の適切な範囲について極性を付与するものもある [Wilson 05]。評価表現の辞書も多くつくられており [Esuli 06]、これらの辞書と機械学習の手法を組み合わせて極性を推定する研究には枚挙にいとまがない。

書き手の感情については、研究が多いわけではないが、近年、ユーモア、皮肉、ヘイトを推定する研究が増えてきたようである。ユーモアや皮肉はテキストそのものから推定することは難しいため、絵文字や Twitter のハッシュタグをメタ情報として用いるなどの工夫がされている [Carvalho 09]。例えば、絵文字がネガティブであり、文章がポジティブであれば皮肉であると推定するといった具合である。

読者の受ける感情を推定する研究はさらに少ないが、対話システムにおいて発話が喚起するユーザの感情を推定する研究 [長谷川 14] や、ユーザ満足度の推定の研究 [Higashinaka 10]、システムの発話が不適切とユーザに思われるかを推定する「対話破綻検出」などの研究 [Higashinaka 15] がある。ユーザの受止め方は多種多様であり、高精度で推定することはそもそも難しい。一般に、ある発話に対するユーザの主観的な評価の一致率は低い (k 値で 0.2 ~ 0.4 程度)。主観的な評価をどのように扱うかは、テキスト処理の観点からのみならず、現在の対話システム全体の大きな課題となっている。

3.2 音声処理

音声と感情に関する研究は、心理学寄りの基礎研究と、工学寄りの応用研究に分けられる。特に、感情と音声の音響的特徴との関係は、このどちらにとっても重要な基盤をなす研究対象である。

感情によって影響を受ける音声の音響的特徴には、韻律的特徴 (F0 変化パターン, 強度変化パターン, 継続時間/話速), 音源由来の声質パラメータ (平均的 F0, 平均的強度, ジッタ, シマ, 喉頭雑音, 声門開放率, スペクトル傾斜), 声道由来の声質パラメータ (フォルマント周波数, フォルマント帯域幅) が含まれる。ただし, 基本感情だけを対象とした場合でさえも, 感情カテゴリーの違いが音響的特徴に及ぼす影響について, 一般的に述べるができる事実はあまりない。一方, 感情次元については, 覚醒-睡眠次元は F0 や強度などの音響的特徴に強く反映されることが広く知られている [森 14]。

工学寄りの研究としては, 音声からの話者の感情を機械に認識させる研究と, 合成音声に感情を付与する研究が代表的である。前者で多いのは, 「感情分類」とでも呼ぶべき研究枠組みである。これは, 文字認識などのパターン分類と同じく, N 種類の中から一つの感情カテゴリーを「正解」として割り当てられた未知の音声サンプルに対し, その音響的特徴から「正解」の感情カテゴリーを推定する問題である。

残念ながら, 感情分類に関する研究論文の大部分は, それ自身が応用上どのような意義をもつのかを明確に述べていない。一方で, 音声からの特定の感情を検出する技術には明確な需要がある。その一つの例が, 顧客の怒り検出 [Pappas 15] である。これは, コールセンタの対話において, システムによる自動応答でコミュニケーションがうまくいかない場合に人間のオペレータにスイッチしたり, 後日対話ログを分析したりするのに利用できる。上述した感情に関連する音響的特徴のほか, 音声に含まれる言語情報が怒り検出に有効であるという研究がある [有本 07, Nomoto 11, Polzehl 11]。

コミュニケーション場面で話し方や姿勢などの非言語行動が似てくる現象は同調 (synchrony, contagion, または, mimicry) と呼ばれる。一般に, 同調の傾向は共感 (emphathy) や社会性などの要因の影響を受ける。特に, 強い同調傾向はポジティブな社会的関係を反映すると考えられている。例えば, 対話相手の発話タイミングが自分と似ていると, 相手を「親しみやすい」, 「好きな」と評価する傾向がある [長岡 05]。このような音声のレベルでの同調現象を実装することで, ユーザの対話システムに対する印象を改善できる可能性がある。例えば, CG キャラクターが人間の発話の韻律をまねてオウム返しすると, より好意的に評価されることを示したものがある [Suzuki 03]。このような感情表出に関連した非言語行動のモデル化は, 人と良い関係を築く対話システ

ムの振舞いを設計するうえでも重要である。

3.3 マルチモーダル処理

コミュニケーションにおける感情状態は, テキストや音声だけでなく, 例えば, 表情・ジェスチャといった視覚的な非言語情報 [Knapp 13] や, 心拍・血圧といった生理指標からの情報にも現れることが広く知られている [Cacioppo 00]。システム側から見れば, ユーザの音声・言語だけでなく, 他の非言語情報を入力情報として, 対話状況を理解したり, ユーザの意図・態度・状態推定を行うことで, より自然な対話を実現できると期待されている [中野 15]。マルチモーダル処理では, 主に, 視覚情報を中心として, テキスト・音声以外のモダリティを使った感情推定と, バーチャルエージェントの感情状態の表出のための感情生成の研究に分かれる。

視覚情報の中で, 最も感情状態が表出するのは表情であるといわれており, 表情の分析・表情からの感情推定に関する研究が盛んに行われている [Zeng 09]。Ekman らの開発した Facial Action Coding System (FACS) [Ekman 97] を用いて, 顔の各部位の動作パターンの組合せとして表情を記述する方法が, 顔画像からの特徴量の記述や感情状態推定に利用されることが多い。表情以外には, 手・体のジェスチャ, 頭部動作, 姿勢変化といった特徴も感情推定に有効であることがわかっている。表情の情報を含めた視覚情報と音声情報を統合したマルチモーダル情報を用いて, 発話者の感情状態の推定精度を向上させるための研究も盛んに行われている [Wu 14]。一般的に, 音声と視覚データを統合することで, 感情推定精度が向上すると述べられており, マルチモーダル情報の統合は有用なアプローチである。

マルチモーダル処理で感情状態の推定を行ううえで, 困難な課題も多い。例えば, 信頼性の高い感情状態のアノテーション方法や, 非言語行動の個人差のモデル化に関する検討・知見は十分ではない。また対話データ収集が, 実験室・実環境で行われた場合の相違や, 表情や感情をシステム (もしくはロボット) に推定されていることを, 実験協力者が未知・既知であるかの相違といった, さまざまな設定環境の違いと, 観測される非言語情報の違いの影響の検証も必要である [Picard 01]。

人型の対話ロボットやエージェントのための非言語行動生成やマルチモーダル情報の生成に関する研究が多く行われている [Gunes 11, Pelachaud 09]。特に, 感情状態の表出には表情の制御に関する方法が研究されており, ここでも FACS が表情表出モデルの基礎となっている。表情以外には, 例えば, 感情的な発話生成を行うための, 頭部動作の生成方法が提案されている [Busso 07]。バーチャルエージェントの研究では, 話し手行動と聞き手行動の生成の両面から研究が進められており, 感情の表出によって, ユーザに「親近感をもたせる」, 「対話する気を起こさせる」, 「信頼感を与える」, 「共感を与

える」ための、マルチモーダル情報の生成方法が目下の研究課題である。

最後に、マルチモーダル情報の推定・生成の利点を生かしたシステムを紹介する。SimSensei と呼ばれる対話エージェントは、リアルタイムにユーザの音声・表情を観測し、ユーザの感情、特に、苦悩、苦痛の状態を推定しながら、問診を行う [DeVault 14]。推定したユーザの苦悩、苦痛の状態を、医師にフィードバックすることで、ストレスを抱えるユーザの治療の支援を行う。

3.4 ロボティクス

人とのインタラクションを無視できない存在としてロボットがある。従来のロボティクスにおける感情の扱いは、対話システムのような言語を使ったインタラクションとは切り離されて議論されている。例えば Miwa らは、大きな音や強い光、高温などのセンサ情報を用いて評価された感情をもとに、興味を示したり嫌悪の表情を表出する機能をもつ頭部ロボットを開発した [Miwa 01]。これは、ある種の自己防衛反応とみなすことができ、生き残るための行動の原動力として感情を捉える進化論学派の考え方に則っているといえよう。また、Brezeal は、後述するアプレイザル理論を引用し、人の声のトーンや顔の動きなどをもとに評価した感情を、表情として出力するロボットを開発した [Brezeal 03]。このロボット自体は言葉をしゃべれない幼児を想定したものであるが、ロボットの感情的なフィードバックによりインタラクションが活発になることを実験により確認している。

このように、ロボティクス分野では、言語コミュニケーションとは関係性が低い中で感情の研究が行われてきた経緯から、ペット型や、人型であっても幼児を模倣したものに言葉をしゃべれないロボットを想定している例が多かった。しかし近年、音声対話機能をもつロボットが一般販売されるなど、ヒューマンロボットインタラクション研究の様相が変わりつつある。なかでも Softbank 社が販売している Pepper が注目されている。このロボットは感情認識ロボットを標榜しており、感情マップというアプリを利用することでロボット内部の感情状態を可視化することができる [Softbank]。また、Pepper のハードウェア開発を担当している Aldebaran 社が開発・販売している Nao も、コミュニケーションの研究に広く利用されている。このように、対話システムを搭載したロボットが身近になってきていることから、今までの研究成果を言語コミュニケーションにいかに関活用していくかがこの分野における次の大きな課題である。

4. 対話システムにおける感情モデル

これまで見てきたように、対話システムに関する各分野においては、テキスト、音声、もしくは、マルチモー

ダル情報を用いて基本感情を推定し、その感情に従って応答を行うものが多いが、感情をもったシステムを実現するには、感情の認識だけではなく生成もできる必要があるだろう。しかしながら、対話システムで感情モデル（感情生起のメカニズム）を内部に持ち合わせたものは少ない。感情モデルを実現するためには、感情がどのように生起するのかについての理論が必要である。ここでは、比較的認知されている「アプレイザル理論」を紹介する。また、アプレイザル理論をベースに提案されている感情モデルと対話システムにおける計算モデルを紹介する。

4.1 アプレイザル理論

アプレイザル理論とは、いくつかの評価次元に沿って環境からの情報を評価した結果として感情を説明するものである [Scherer 99]。すなわち、環境の変化があると、まずそれがどのようなものを評価し、その評価に従って、感情が生起すると考える。この理論の良いところは、同じ現象について、個人ごとに異なる感情をもち得ることの説明ができるということである。

アプレイザル理論の中にもさまざまなあるが、後述する計算モデルの EMA で用いられている Lazarus の考え方を紹介すると、評価の段階として、primary appraisal と secondary appraisal がある [土田 96]。Primary appraisal では、環境において生起した出来事そのものの自分との関係性や望ましさなどが評価される。Secondary appraisal では、その出来事に対してどのように自分が対処できるか（コーピングと呼ぶ）の観点から評価される。コーピングには 2 種類あり、problem-focused coping と emotion-focused coping がある。前者は、出来事への直接的な対処に関するもので（例えば、起こった問題を解決する）、後者は出来事への心理面での対処に関するものである（例えば、問題がなかったことと考える）。これらの 2 種類の評価に基づき、最終的な感情が決定される。例えば、望ましくなく、対処できないような出来事であると評価されたら「悲しみ」の感情が生起するといった具合である。また、同様の出来事が、その責任が他人にある場合は「怒り」となったりする。

4.2 Component Process Model

Component Process Model [Scherer 84] は、感情状態の違いを、環境の変化に対して生体システムが実行する一連の評価（刺激評価チェック）の結果だと考えるモデルである。刺激評価チェックは、(1) 新規性、(2) 快、(3) 重要性、(4) コーピング可能性、(5) 社会規範や自己適合性からなり、これらのチェックの結果が自律神経系に特定のパターンの反応を引き起こす。Scherer は、この仮説を感情音声の音響特徴の予測に用いた [Scherer 86]。例えば、恐れに関連する刺激からは新規性高、不快、重要性高、対処可能性低などの評価が得られ、各刺

激評価チェックから予想される生理的反応と組み合わせることで、F0 揺らぎ、F0 平均、F0 レンジ、F0 標準偏差、高域周波数エネルギー、話速の増大を予想することができている。

Scherer により開発された GENESE [Scherer 93] は、アプレイザル理論に基づいた感情の計算モデルを実装した初期のシステムである。強い感情を経験した状況に関する情報を入力すると、その感情状態を予測して出力する。各感情に対する知識ベースは、その感情に対する典型的な刺激評価チェックの出力の予測値を要素としてもつベクトルであり、これらは Scherer 自身による過去の研究から編み出されたものである。ユーザは、コンピュータから呈示される 15 の質問に対して多肢選択で回答する。これらの質問は、ある特定の刺激評価チェックに相当する。例えば、質問 1「あなたの感情を喚起した状況は、突然起こったものですか？」は新規性チェックに対応する。実験の結果、253 の感情エピソードに対し、14 種類のカテゴリの中から 77.9% の割合で正しい感情を推定することができたと報告されている。

4.3 EMA (Emotion and Adaptation)

EMA は [Smith 90] の考え方を実装した計算モデルである [Gratch 04]。EMA では、環境の評価値 (appraisal variable) として、関係性、望ましさ、出来事の責任の所在、出来事の起こりやすさ、予期できたかどうか、緊急性、自己との関わり、コーピング可能性などをもつ。

GENESE では、これらの評価値をユーザが入力していたが、対話システムとの会話中にユーザが個々の評価値を入力することは難しい。よって、これらを自動的に推定する機構が必要となる。EMA では、これらの評価値を自動的に導出するために、決定理論的プランニングを用いる。決定理論的プランニングを用いることで、不確実性のある中、因果関係の推論によって、関係性や望ましさを計算することができる。システムは、自身の信念、欲求、意図をもち、現在の状況に従ってどのように行動すると、どのような結果が得られるかを逐次計算する。システムは、評価値に基づいて感情 (いわゆる基本感情に即したもの) を表出し、コーピング可能性が高く、状況を改善することができるアクションをコーピングとして実施する。システムがコーピングを行うと環境が変わるため、新しい環境に基づき評価が行われる。

EMA は、バーチャルヒューマン研究の感情モデルとして実際に用いられており、現状対話システムにおいて最も受け入れられているモデルである。ただ、実際に用いられていると言っても、対話におけるプランニングの難しさから適用範囲は限定的である。また、評価値から感情へのマッピングやコーピングのためのアクションの選択などにヒューリスティックが多く用いられており、これらの自動設定も今後必要だろう。

なお、EMA 以外にも、感情の計算モデルはいくつか

存在している。例えば、Becker らは、アプレイザル理論に基づく感情生起モデルと PAD モデルの感情次元に基づく行動生成モデルを統合した WASABI を提案している [Becker-Asano 08]。また、PEACTIDM [Marinier 09] など、多くのモデルはアプレイザル理論をもとに構築されている。感情モデルに関して興味のある方は、[Marsella 10] を参照されたい。

5. おわりに

本稿では、対話システムと感情に関する研究をサーベ的に紹介した。ここまで述べてきたように、アプレイザル理論などの動的な感情生起モデルを陽に組み込んだ対話システムの研究はまだ十分に行われていない。EMA などの感情の計算モデルと、現在までに培われてきた対話システムの技術を統合することで、より人間との親和性の高い対話システムが構築できるだろう。これからの対話システムには感情モデルが搭載され、ユーザと心が通い合うような対話が実現されていくことを期待している。

謝辞

大変有益なコメントをいただきました宇都宮大学国際学部中村 真先生、ならびに成蹊大学理工学部中野有紀子先生に感謝いたします。

◇ 参考文献 ◇

- [有本 07] 有本泰子, 大野澄雄, 飯田 仁: 「怒り」の発話を対象とした話者の感情の程度推定法, 自然言語処理, Vol. 14, No. 3, pp.147-163 (2007)
- [Becker-Asano 08] Becker-Asano, C.: WASABI: Affect Simulation for Agents with Believable Interactivity, Faculty of Technology, University of Bielefeld, IOS Press (DISKI 319) (2008)
- [Breazeal 03] Breazeal, C.: Emotion and sociable humanoid robots, *Int. J. Human-Computer Studies*, Vol. 59, pp. 119-155 (2003)
- [Busso 07] Busso, C., Deng, Z., Grimm, M., Neumann, U. and Narayanan, S.: Rigid head motion in expressive speech animation: Analysis and synthesis, *IEEE Trans. on Audio, Speech, and Language Processing*, Vol. 15, No. 3, pp. 1075-1086 (2007)
- [Cacioppo 00] Cacioppo, J. T., Berntson, G. G., Larsen, J. T., Poehlmann, K. M. and Ito, T. A., et al.: The psychophysiology of emotion, *Handbook of Emotions*, Vol. 2, pp. 173-191 (2000)
- [Carvalho 09] Carvalho, P., Sarmiento, L., Silva, M. J. and De Oliveira, E.: Clues for detecting irony in user-generated contents: oh...!! it's so easy;-), *Proc. CIKM Workshop on Topic-Sentiment Analysis for Mass Opinion*, pp. 53-56 (2009)
- [DeVault 14] DeVault, D., Artstein, R., Benn, G., Dey, T., Fast, E., Gainer, A., Georgila, K., Gratch, J., Hartholt, A. and Lhommet, M., et al.: SimSensei Kiosk: A virtual human interviewer for healthcare decision support, *Proc. AAMAS*, pp. 1061-1068 (2014)
- [Ekman 72] Ekman, P.: Universals and cultural differences in facial expressions of emotion, in Cole, J., ed., *Nebraska Symp. on Motivation*, pp. 207-283, University of Nebraska Press, Lincoln (1972)
- [Ekman 97] Ekman, P. and Rosenberg, E. L.: *What the Face*

- Reveals: Basic and Applied Studies of Spontaneous Expression Using the Facial Action Coding System (FACS)*, Oxford University Press (1997)
- [Esuli 06] Esuli, A. and Sebastiani, F.: Sentiwordnet: A publicly available lexical resource for opinion mining, *Proc. LREC*, Vol. 6, pp. 417-422 (2006)
- [Gratch 04] Gratch, J. and Marsella, S.: A domain-independent framework for modeling emotion, *Cognitive Systems Research*, Vol. 5, No. 4, pp. 269-306 (2004)
- [Gunes 11] Gunes, H., Schuller, B., Pantic, M. and Cowie, R.: Emotion representation, analysis and synthesis in continuous space: A survey, *Proc. FG and Workshops*, pp. 827-834 (2011)
- [長谷川 14] 長谷川貴之, 鍛冶伸裕, 吉永直樹, 豊田正史: オンライン上の対話における聞き手の感情の予測と喚起, *人工知能学会論文誌*, Vol. 29, No. 1, pp. 90-99 (2014)
- [Hatfield 93] Hatfield, E., Cacioppo, J. and Rapson, R.: Emotional contagion, *Current Directions in Psychological Science*, Vol. 2, pp. 96-99 (1993)
- [Higashinaka 10] Higashinaka, R., Minami, Y., Dohsaka, K. and Meguro, T.: Issues in predicting user satisfaction transitions in dialogues: Individual differences, evaluation criteria, and prediction models, *Spoken Dialogue Systems for Ambient Environments*, pp. 48-60, Springer (2010)
- [Higashinaka 15] Higashinaka, R., Funakoshi, K., Araki, M., Tsukahara, H., Kobayashi, Y. and Mizukami, M.: Towards taxonomy of errors in chat-oriented dialogue systems, *Proc. SIGDIAL*, pp. 87-95 (2015)
- [Knapp 13] Knapp, M., Hall, J. and Horgan, T.: *Nonverbal Communication in Human Interaction*, Cengage Learning (2013)
- [Marinier 09] Marinier, R. P., Laird, J. E. and Lewis, R. L.: A computational unification of cognitive behavior and emotion, *Cognitive Systems Research*, Vol. 10, No. 1, pp. 48-69 (2009)
- [Marsella 10] Marsella, S., Gratch, J. and Petta, P.: Computational models of emotion, in Scherer, K. R., Bänziger, T. and Roesch, E. B., eds., *Blueprint for Affective Computing — A Sourcebook*, pp. 21-41, Oxford University Press, Oxford (2010)
- [Mehrabian 74] Mehrabian, A. and Russell, J. A.: *An Approach to Environmental Psychology*, MIT Press (1974)
- [Minsky 07] Minsky, M.: *The Emotion Machine: Commonsense Thinking, Artificial Intelligence, and the Future of the Human Mind*, Simon & Schuster (2007)
- [Miwa 01] Miwa, H., Umetsu, T., Takanishi, A. and Takanobu, H.: Robot Personality based on the Equations of Emotion defined in the 3D Mental Space, *Proc. ICRA*, pp. 2602-2607 (2001)
- [森 14] 森 大毅, 前川喜久雄, 粕谷英樹: 音声は何を伝えているか—感情・パラ言語情報・個人性の音声科学—, コロナ社 (2014)
- [長岡 05] 長岡千賀, 小森政嗣, 中村敏枝: 音声対話における反応潜時が話者印象評定に及ぼす影響: 社会的スキルの程度による評定の手がかりの相違, *信学技報, HCS2004-70* (2005)
- [中野 15] 中野幹生, 駒谷和範, 船越孝太郎, 中野有紀子: 対話システム (自然言語処理シリーズ), コロナ社 (2015)
- [Nomoto 11] Nomoto, N., Tamoto, M. and Masataki, H.: Anger recognition in spoken dialog using linguistic and paralinguistic information, *Proc. Interspeech*, pp. 1545-1548 (2011)
- [Pang 05] Pang, B. and Lee, L.: Seeing stars: Exploiting class relationships for sentiment categorization with respect to rating scales, *Proc. ACL*, pp. 115-124 (2005)
- [Pappas 15] Pappas, D., Androutsopoulos, I. and Papageorgiou, H.: Anger detection in call center dialogues, *Proc. IEEE CogInfoCom*, pp. 139-144 (2015)
- [Pelachaud 09] Pelachaud, C.: Modelling multimodal expression of emotion in a virtual agent, *Philosophical Trans. of the Royal Society B: Biological Sciences*, Vol. 364, No. 1535, pp. 3539-3548 (2009)
- [Picard 01] Picard, R. W., Vyzas, E. and Healey, J.: Toward machine emotional intelligence: Analysis of affective physiological state, *IEEE Trans. on Pattern Analysis and Machine Intelligence*, Vol. 23, No. 10, pp. 1175-1191 (2001)
- [Polzehl 11] Polzehl, T., Schmitt, A., Metze, F. and Wagner, M.: Anger recognition in speech using acoustic and linguistic cues, *Speech Communication*, Vol. 53, No. 9-10, pp. 1198-1209 (2011)
- [Popescu 07] Popescu, A.-M. and Etzioni, O.: Extracting product features and opinions from reviews, *Natural Language Processing and Text Mining*, pp. 9-28, Springer (2007)
- [Russell 80] Russell, J. A.: A circumplex model of affect, *J. Personality and Social Psychology*, Vol. 39, No. 6, pp. 1161-1178 (1980)
- [Scherer 84] Scherer, K. R.: Emotion as a multicomponent process: A model and some cross-cultural data, in Shaver, P., ed., *Review of Personality and Social Psychology*, Vol. 5, pp. 37-63, Sage Publications, Beverly Hills, CA (1984)
- [Scherer 86] Scherer, K. R.: Vocal affect expression: A review and a model for future research, *Psychological Bulletin*, Vol. 99, No. 2, pp. 143-165 (1986)
- [Scherer 93] Scherer, K. R.: Studying the emotion-antecedent appraisal process: An expert system approach, *Cognition and Emotion*, Vol. 7, No. 3-4, pp. 325-355 (1993)
- [Scherer 99] Scherer, K. R.: Appraisal theory, in Dagleish, T. and Power, M., eds., *Handbook of Cognition and Emotion*, pp. 637-663, Wiley (1999)
- [Smith 90] Smith, C. A. and Lazarus, R. S.: *Emotion and Adaptation*, Guilford Press (1990)
- [Softbank] <http://www.softbank.jp/robot/>
- [Suzuki 03] Suzuki, N., Takeuchi, Y., Ishii, K. and Okada, M.: Effects of echoic mimicry using hummed sounds on human-computer interaction, *Speech Communication*, Vol. 40, No. 4, pp. 559-573 (2003)
- [土田 96] 土田昭司, 竹村和久: 感情と行動・認知・生理—感情の社会心理学, 誠信書房 (1996)
- [Turney 02] Turney, P. D.: Thumbs up or thumbs down?: Semantic orientation applied to unsupervised classification of reviews, *Proc. ACL*, pp. 417-424 (2002)
- [Wilson 05] Wilson, T., Wiebe, J. and Hoffmann, P.: Recognizing contextual polarity in phrase-level sentiment analysis, *Proc. HLT/EMNLP*, pp. 347-354 (2005)
- [Wu 14] Wu, C.-H., Lin, J.-C., and Wei, W.-L.: Survey on audiovisual emotion recognition: databases, features, and data fusion strategies, *APSIPA Trans. on Signal and Information Processing*, Vol. 3, p. e12 (2014)
- [Zeng 09] Zeng, Z., Pantic, M., Roisman, G. I. and Huang, T. S.: A survey of affect recognition methods: Audio, visual, and spontaneous expressions, *IEEE Trans. on Pattern Analysis and Machine Intelligence*, Vol. 31, No. 1, pp. 39-58 (2009)

2016年8月2日 受理

著者紹介



東中 竜一郎 (正会員)

1999年慶應義塾大学環境情報学部卒業。2001年同大学院政策・メディア研究科修士課程。2008年博士課程修了。2001年日本電信電話株式会社入社。現在、NTTメディアインテリジェンス研究所に所属。質問応答システム・音声対話システムの研究開発に従事。博士(学術)。言語処理学会、情報処理学会、電子情報通信学会各会員。



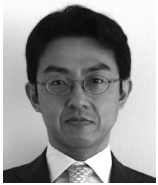
岡田 将吾 (正会員)

2003年横浜国立大学工学部卒業。2008年東京工業大学大学院総合理工学研究科博士課程修了。同年、京都大学情報学研究科特定助教。2011年東京工業大学総合理工学研究科助教。2014年IDIAP Research Institute 滞在研究員。2016年東京工業大学情報理工学院助教。博士(工学)。ACM、電子情報通信学会各会員。マルチモーダルインタラクション、人間行動解析、社会的信号処理の研究に従事。



藤江 真也 (正会員)

1999年早稲田大学理工学部卒業。2001年同大学院理工学研究科修士課程修了。2004年同博士課程単位取得退学。早稲田大学理工学部助手、同大学理工学総合研究所客員講師、同大学高等研究所助教、アズビル株式会社勤務を経て、2014年より千葉工業大学先進工学部未来ロボティクス学科准教授。博士(工学)。専門は音声対話システム・会話ロボット・マルチモーダル情報処理。電子情報通信学会、情報処理学会、日本音響学会、日本感性工学会各会員。



森 大毅

1993年東北大学工学部通信工学科卒業。1998年同大学院工学研究科博士後期課程修了。博士(工学)。同年、東北大学大学院工学研究科助手。現在、宇都宮大学大学院工学研究科准教授。音声合成、音声対話、医療福祉工学の研究に従事。電子情報通信学会、日本音響学会、情報処理学会、ISCA各会員。