

表情音声心情識別



基礎フェーズ 応用フェーズ 実用化フェーズ

広島市立大学 大学院情報科学研究科 目良 和也

ここに注目!! 機械学習で表情や声音から感情を推定し、言葉の裏の本音を探る

表情・声のトーン・発言の内容を統合して判断することで、感情を推定するアルゴリズム。機械学習で感情を推定するシステムの土台が完成。一般的なパソコン、パソコンに搭載されたカメラは、マイクは指向性が高いものを別途用意。

[用途] 医療・介護における会話、モニター調査（商品、サービス等）、電話受付など。



「このお礼は次回、かならず」という言葉。

素直に受け取れば礼儀正しい態度ですが、これがサスペンスドラマの台詞だったとしたら、どうでしょうか。まったく逆の緊迫した場面が浮かんできます。私たちの日常でも、言葉にまったく逆の感情が潜んでいることは少なくありません。

私たちは、発言を、言葉の内容（テキスト）、声のトーン、表情の3要素にわけて、それぞれが示す感情を機械学習により推定するシステムの開発を進めています※1。

表情・音声・テキストが示す感情を同時並行に推定し、それぞれの結果を比較することで、矛盾があれば嘘や隠しごと、お世辞などがあると評価することができます。逆に、示された感情が微妙であっても、それぞれの結果に矛盾がなければ、確度の高い推定結果だと考えることもできます。

2020年は教育や企業活動におけるコミュニケーションの多くが、ネットを介してリモートで行われることになりました。対面コミュニケーションと比較するとリモートでは欠落する情報があると考えられますが、ここに私たちの技術を活用することで欠落部分を補う可能性も検討中です。



※1:構築したデモンストレーションシステム

実用化レベル

表情、音響（声のトーン）、テキスト（発言の内容）の、それぞれについて、機械学習にて感情を推定するシステムの土台は完成しています※2。汎用性を求めると精度が落ちるため、用途に沿った学習を行ってカスタマイズでシステムを構築することが望ましいと考えています。

表情、音響、テキストの3システムを統合して判断するシステムも開発を進めています。

活用シーン

- コールセンターやホットライン（音声）
- 対面での相談、診察（音声、音声+表情）
- モニター評価（商品、サービス開発など）
- 医療、介護（感情表出が困難な患者との意思疎通など）
- 日常コミュニケーションのサポート

技術概要

《表情からの感情推定》表情を捉えた映像を10ミリ秒ごとに静止画として切り出し、いくつかの顔の特徴点から5種類の感情を推定します（無・喜・驚・怒・悲）。連続的な表情の変化を捉えることも、一瞬だけ眉をひそめるなどの微細な変化を捉えることも可能です※3。

《音声からの感情推定》発言の音声情報から周波数、音のエネルギーなどの音響的特徴を抽出し、11種類の感情を推定します（嬉しい・楽しい・リラックス・冷静・不安・怒り・恐れ・嫌悪・軽蔑・疲れた・悲しい）※4。

《テキストからの感情推定》発言内容を音声認識によってテキスト化し、文字列全体が表す感情を3種類に分類します（快・中立・不快）※5。

いずれも感情推定のルールは機械学習によって構築していますが、学習に用いるデータによる制約や話者の性格などによる個人差はありますが、人が他者の発言から感情を推定するのと同程度の推定は可能であることを確認しています。表情・音声・テキストそれぞれからの感情推定だけでなく、それぞれの感情推定を同時に行って補完・強化すると同時に、情報が食い違う場合に隠された本心を見抜くツールになり得ると考えています。

必要機器

機械学習には能力の高い計算機が必要ですが、機械学習によって得られたルールの適用は一般的なパソコンで実行できます。カメラはパソコンに搭載されたもので充分ですが、マイクは機器搭載のものよりも指向性が高いものを別途用意してください。

出力データ

表情から5種類、音声から11種類、テキストから3種類の感情を推定して出力します。推定結果を評価するシステム側で、似た感情を例えば、positive、neutral、negativeの3種類に統合して表示するなどの処理を行っています。

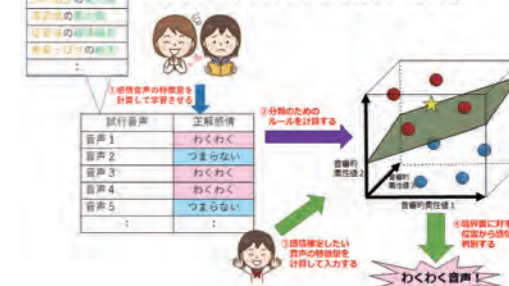
測定方法

現行のシステムでは、起動後にカメラ、マイクに向かって発言を行うことで測定できます。用途に応じて、測定前に性格タイプをアンケートなどで分類するなどの処理が必要になると考えています。

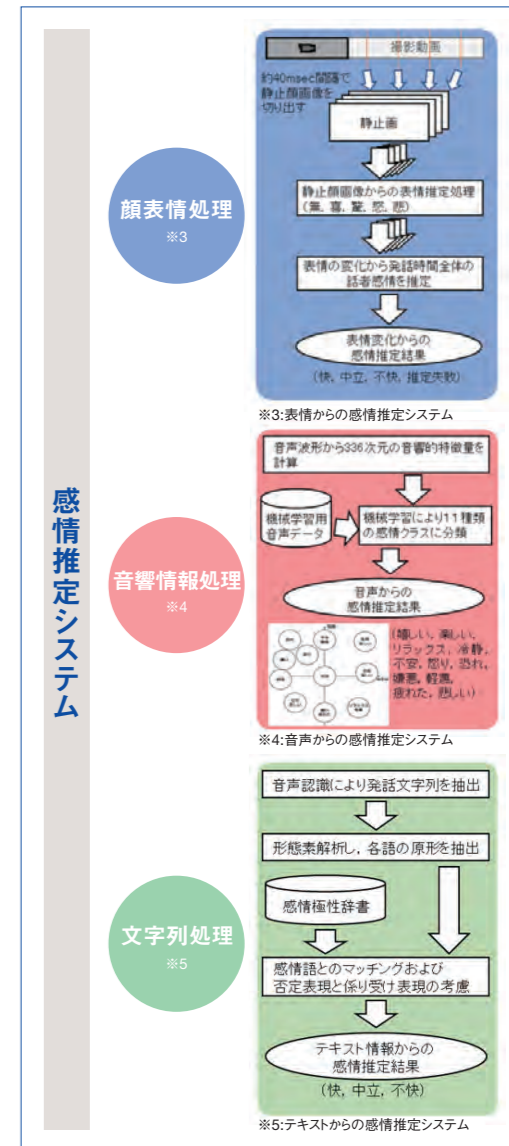
推奨したい他ツールとの統合解析※6

- 感性メーター
- 血管剛性測定
- 顔カメラ・顔血流心拍

機械学習器を用いて口調から感情を読む技術 (SVM)



※2:機械学習とは、与えられたたくさんの正解や間違いの事例から、どのように判断すれば良いかを自分で学習



※6:表情音声心情識別ツールと感性メーターの統合解析実験

社会実装に向けた取り組み

●学習データの検討と収集^{※7}

表情・音声・テキストをもとに話者の感情推定を行うシステムを構築するにあたって、私たちは、「一般的な人間と同等のレベル」、「一般的な人間では見抜けないレベル」の2つの目標を設定しています。ここで課題となるのは、機械学習に用いるデータです。

人間と同等レベルの推定を行うための素材としては、会話を撮影した動画に第三者が推定した感情をラベリングして学習に用いています。

人間を超えるレベルの推定のためには、ゲーム中の会話やフリートークの動画などに、話者本人が会話中の感情を思い出してラベルを付けたデータを収集しています。

また、発言と本心が矛盾する状況のデータのひとつとして、学生が担当教官をひたすら褒める（お世辞を言う）状況を記録しました。ただ、微かな表情の変化や言葉と矛盾する感情をストレスや作為のない状態で記録するのは簡単ではありません。

今後、民間企業や医療・介護などで実用するにあたっては、それぞれの現場の記録データを、許可を得たうえで機械学習に使わせていただくことも考えています。

●コールセンターへの実装を想定した実験

面識のない相手と音声通話で込み入った会話を行うコールセンターのようなシチュエーションで、通話相手の感情を推定する機械学習システムの構築を試みました。

コールセンターでは、顧客がネガティブな感情を抱いた状態で会話がスタートすることも多く、顧客の感情が好転しているのか悪化しているのかを早めに察知したいというニーズがあります。

学習の素材として用いたのはインターネットで収集した動画です。演技などではない素の状態で感情が変化する過程が撮影されたものを選びました。第三者が動画の登場人物が最終的に示す感情をラベルし、機械学習では感情が明確に表出する前の音声を用いたところ、それなりの確率で隠された感情を正しく推定することができました。

社会実装にあたっては、より現場に近いデータを用いることで有効なシステムを構築できると期待しています。

今後の応用可能性

STEP 1 隠された感情を読む

10ミリ秒単位の測定で一瞬の表情変化を捉えたり、音響測定により声のわずかな震えを捉えたり、人間同士では気づけない本心を推定できるシステムの構築をめざしています。

[商品やサービスの評価に]

モニター調査などで製品・サービスについて評価を求められた時、批判的な感想を忌憚なく語れる人はむしろ少数派でしょう。インタビュー時の表情や音声を記録し、解析することで、本心からの評価なのか遠慮が含まれているかを判断することができます。



※7:彼らの言葉は本心、それとも嘘?感情推定の実験風景。



↑コールセンターでの活用(イメージ)



↑インターネット上の動画を用いた機械学習(イメージ)



↑モニター調査をより正確に(イメージ)

[コールセンターなどで]

質問や苦情を受けることの多いコールセンターで、音声から推定できる相手の感情をモニタリングすることで、スタッフが言葉遣いや対応方法を変えたり、新人から先輩にバトタッチするなどの対応が可能になります。顧客満足度の向上、スタッフのストレスの低減に繋がると考えています。

[医療・メンタルヘルスで]

Withコロナの時代、診察や相談事業をリモートで行うケースも増えていくでしょう。同じ空間にいれば察知できるような感情はもちろん、パソコンなどを介して会話している状況を利用して、よりきめ細かな感情推定を行うことで診察・アドバイスに役立てることができそうです。

リモートだけでなく直接の面談でも、推定結果を治療に役立てたり、患者さんひとりひとりにチューンナップした推定を行うことも可能になると考えています。

[犯罪捜査にも]

捜査対象者が本当のことを語っているか否か捜査員の勤に頼っていた部分を、この感情推定システムによって補うことができるのではないかと考えます。特に、表情や音声の微細な変化を読み取ることで、嘘をつき慣れた人の本心を知ることができるかもしれません。

STEP 2 人と同レベルで感情を読む

人が読めるレベルの感情を推定するシステムにも需要はあります。たとえば、日々のメンタルヘルスチェック、アバターをつかったビデオ通話、AIとの対話などの場面です。

[メンタルヘルスチェック]

スマートフォンの健康アプリと対話することで表情変化の速度や声の調子を記録します。不調の徴候を発見した場合、気分転換や受診を提案することができます。

[AIやサポートロボットとの対話]

質問を受け付けるチャットbotや、受付、給仕を行うサービスロボットなどが増えています。

そこに客の感情を推定するシステムを導入することで、より円滑なサービスを実現できます。

[リモート対話のサポート]

ネットを利用したビデオ会議が当たり前になりましたが、顔を映したくない状況も珍しくありません。通話アプリに表情推定システムを組み込み、通話者の代わりに実際の表情を反映したアバターを表示するという活用が考えられます。

また、リモート授業など1対多数となる通話において生徒の感情を表示するというニーズもあるでしょう。ライブ配信で視聴者の反応を演者にフィードバックするという活用方法もありえるかもしれません。

[人の代わりに空気を読む]

自閉症スペクトラム障害などで、言葉を額面通りに受け取ってしまったり、空気を読めないために生きづらさを感じている人たちがいます。スマートフォンなどに対話相手の感情の推定するアプリを搭載することで、生活をサポートできるのではないかと期待を持っています。



↑遠隔医療(カウンセリング)での活用(イメージ)



↑犯罪捜査(取り調べ)での応用(イメージ)



↑スマートフォンやウェアラブルデバイスを用いて簡単にメンタルヘルスチェック(イメージ)



↑ロボットと心を通わせられるかも?(イメージ)



↑リモート会話を、より「気持ち伝わる」コミュニケーションに