

情報科学特論 C レポート

M20A25 中塚淳平

課題 1：感情認識アルゴリズムでは、両者の結果の積集合をとることで結果の結合を行った。それはなぜか。

音声対話システムとは、音声認識、音声合成や音声対話といった技術を駆使することによって、人間と機械が自然に対話可能となることを目指すシステムである。このシステムが完成することで、自然なインターフェイス設計や機械が人間の感情を察することが可能となる。しかし、現在の音声対話システムの認識率は悪い。そのため、事前に準備された対応をとることしかできないので、理想の音声対話システムの完成はまだ遠いというのが現実である。そこで、近年は「深層学習を用いた認識率のさらなる向上」、「雑音抑制」、「音声区間検出」や「発話者の意図の理解」などといった研究が盛んにおこなわれている。

本講義では、「音声認識による勘定識別」に注目していた。従来は音声の特徴量（“声の高さ”，“声の強さ”，“スペクトル形状”など）の時系列を抽出し，その各時系列の統計量（“平均”，“分散”，“最大値”，“線形回帰係数”など）を計算することで，音声から感情を識別していた。従来の感情識別法についてまとめたものを図 1 に示す。

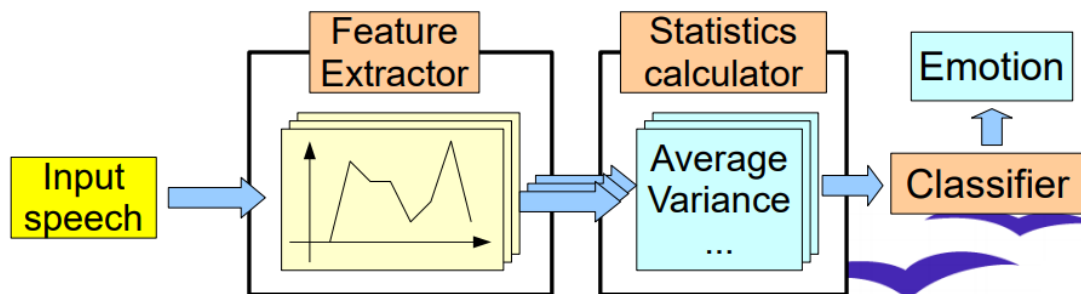


図 1 従来の感情識別法

しかし、従来法には問題点がある。それは、音声の韻律的特徴（声の大きさや高さ）は感情に関係するが 1 つの感情に絞れないという点である。例として、声の高さが高いかつ声が大きく早口で発話しているという特徴において、発話者が嬉しくて興奮しているのか怒って興奮しているかを識別することはできない。

そこで、「発話内容に注目」することで発話内容自体から感情の推定を目指した。しかし、時と場合によって、全く違う感情が正解となる場合が存在する。

よって、「音響的な特徴量」と「発話内容」の両方の特徴量を併用することで音声からの感情認識を目指した。これによって、「特徴量レベルでの統合」と「推定結果レベルでの統

合」の2種類の併用法ができた。特徴量レベルでの統合は、まず2種類の特徴量ベクトルを連結することで、1つのベクトルを作成する。次に、それを識別機に入力することによって感情を識別する手法である。それに対して、推定結果レベルでの統合では、2種類の特徴量をそれぞれ独立に識別を行い、その結果を統合することによって感情を識別する手法である。そのため、推定結果レベルでの統合のほうが推定アルゴリズムや推定モデルなども独立して作成することが可能なため、自由度が非常に高い。しかし、発話内容のみで感情推定を行うため、間違えたスコアを統合する可能性がある。

音響的な特徴量からの感情推定は、前述したとおり推定感情を1つに絞れないという問題点があった。しかし、候補中に正解感情も存在する。それに対して、発話内容からの感情推定は、一つの感情を推定可能だが、全く違う感情を推定する可能性があるという問題点があった。そこで、この2種類の結果から積集合を用いて判定し、積集合が存在。つまり、識別した感情が一致しているならば、その感情とする。また、積集合が空。つまり、2種類の識別結果が合致しない場合、間違えているのは発話内容から推定した結果であるといえるため、音声特徴量からの識別結果のみから感情を識別する。

・課題2：音声から感情が100%の精度で認識できた場合、自然な対話システム以外にどのような応用例が考えられるか。

音声からの感情認識が100%成功するということは、対話内容における感情の起伏を完璧に把握することが可能であるといえる。そこで、「最強のウソ発見器システム」を提案する。感情の起伏を完璧に把握することが可能なので、嘘をついているときの焦りや緊張を判別することが可能である。このシステムが開発されることによって、警察機関での取り調べや裁判などといった、真実のみを発言することが求められる場で活躍することが期待できる。また、スキャンダルや偏向報道がたびたび話題となるマスメディアでの活用も期待できる。