

< 修 士 論 文 >

コールセンターの応対音声品質の  
自動評価に向けた要素技術の研究  
(要 旨)

滋 賀 大 学 大 学 院  
デ ー タ サ イ エ ン ス 研 究 科  
デ ー タ サ イ エ ン ス 専 攻

修了年度：2020年度

学籍番号：6019111

氏 名：高山 和明

指導教員：市川 治

提出年月日：2021年1月20日

コールセンターを運営する企業では顧客対応の中心となる優秀なオペレーターの確保及び育成、各種業務の効率化が喫緊の課題となっている。これらの課題に対して機械学習技術を含む IT 技術の活用は有効な解決策となりうる。例えば現在人手で行われているオペレーターの対応品質評価の自動化である。それによりオペレーターに対する適切な改善指導の実現、管理者が評価作業を効率化しオペレーターへの支援・指導を充実させる事によるサービス品質向上、コスト削減など様々なメリットが期待できる。そのためコールセンターを運営する企業や IT 各社はこの課題に精力的に取り組んでいる。実用化されているものの例では、発話速度、発話かぶりの回数や時間、必須ワードや NG ワードの使用回数などを通話録音データから抽出し、それに対してコンピューターによる自動評価が行われている。これらは主として音声データから言語解析型の音声認識技術によってテキスト情報を抽出し、音声とテキストを対応づける事によって実現されていると考えられる。しかし、「オペレーターが通話相手に良い印象を与える声の表情で話しているか」といった音響的な評価指標に対する自動評価が行われているものは現時点では見当たらない。

本研究では、コールセンターの業務の中で未だ自動化が行われていない「声の表情」の評価の自動化を目標として、その実現可能性の検討を行った。「声の表情」の評価が高い事は「適切な感情を声にのせて表現できている事」と同義であると捉え、音響解析型の音声感情認識の技術を使用して「声の表情」の自動評価を試みた。

まず予備実験として既存手法を用いて音声感情認識を行い、正解率、再現率、適合率を指標として推定精度を確認した。「喜び」「怒り」「悲しみ」「平静」の4感情を目的変数として教師あり機械学習によって分類器を作成した。機械学習への入力とする特徴量には先行研究で広く用いられている JTES (Japanese Twitter-based Emotional Speech) 感情音声コーパスの音声データから抽出した IS10 (INTERSPEECH 2010 Paralinguistic Challenge) 音響特徴量セットを使用した。その結果、先行研究と同程度の推定精度を得られる事が確認できた。次にコールセンターの電話回線の通話品質を想定して JTES の音声データのサンプリング周波数を下げて音質を劣化させたのちに音声感情認識を行い、推定精度の変化を確認した。その結果、正解率、再現率、適合率がそれぞれ 2%程度低下した。この推定精度の低下

を補うために、提案法として発話末尾 0.5 秒分の音声データから抽出した IS10 音響特徴量を特徴量ベクトルに追加して音声感情認識を行い、推定精度の変化を検証した。その結果、正解率が 3%程度向上し、音質の劣化による推定精度の低下を提案法によって補える事が確認できた。

次にコールセンターの音声データを使用して「声の表情」の評価の自動化を試みた。目的変数は人手で付与された「1」「2」「3」の3段階の評点とした。提案法として次の2つを検証した。提案法①：オペレーターが発話した音声から IS10 音響特徴量を抽出し、人手で付与された「声の表情」の評点を正解ラベルとして機械学習を行う。提案法②：話者依存性のある IS10 音響特徴量を話者依存性の少ない特徴量に変換し、IS10 音響特徴量とあわせて機械学習への入力とする。実験の結果、提案法①では評点「2」の音声について適合率 0.77、再現率 0.83 となり、比較的高い精度で推定できた。評点「3」の音声については適合率 0.52、再現率 0.42 となり、完全にランダムな推定を行った場合の適合率 (0.27) と再現率 (0.33) より高い数値となった。また、評点「3」で推定結果が外れたケースの大半は評点「2」から「3」への誤分類、あるいは評点「3」から「2」への誤分類であり、評点「1」から「3」への誤分類と評点「3」から「1」への誤分類は少数であった。つまり評定「3」に関しては推定結果を大きく外したものは少ないため、今後の改善に向けて希望が持てる結果である事が確認できた。評点「1」については適合率 0.19、再現率 0.18 となり、自動化に向けて課題の残る結果となった。提案法②では評点「1」の適合率が 0.29 に向上し、それ以外の指標は提案法①とほぼ同じであった。話者依存性が少なくなるように変換した特徴量を用いる事は「声の表情」の推定精度を上げる効果がある事が確認できた。

まとめとして、本研究ではコールセンターの応対品質評価の中でこれまでに実現していない「声の表情」の評価の自動化について実現可能性を示すことができた。音声感情認識で用いられる特徴量と本研究の提案法により話者依存性が少なくなるように変換された特徴量は「声の表情」の評価の際に有効な特徴量である事が確認できた。今後の課題として、人による評価作業を置き換えるのに十分な精度を得るために評点「1」および「3」の音声の特徴を更に捉える音響特徴量を検討していく必要があると考えられる。