
研 究 紹 介

罰金付きロジスティック回帰マシンによる話者識別

統計計算開発センター 助教授 松井 知子

話者識別は、入力音声为谁の声であるかを判定する技術である。予めシステムに複数の話者の声を登録しておき、入力音声とそれらの話者との類似度を計算して、最大類似度を示す話者を判定結果とする。その際、話者の声をいかに登録するか、いかに入力音声との類似度を計算するかが課題となる。一般に、混合ガウス分布モデル (Gaussian Mixture Model; GMM) を用いて、話者ごとにその学習用音声の特徴ベクトルの分布を表し、類似度としては入力音声に対する GMM の尤度を利用することが多い。GMM パラメータの推定には最尤推定法や最大事後確率推定法が用いられる。それらのモデルごとの推定量は、学習データが十分にある場合や事前分布が適切に設定される場合には高い識別性能を示す。一方、識別という目的を考えた場合、上記の話者ごとの分布自体を効果的に表そうとするアプローチとは異なり、全話者のデータを一度に扱って、各話者の分布境界を識別的に求めるアプローチも考えられる。

現在、各話者の分布境界を識別的に求めるアプローチについて、田邊國士教授 (予測制御研究系) と共同で検討している。具体的には、罰金付きロジスティック回帰マシン (dual Penalized Logistic Regression Machine; dPLRM) [1] の話者識別への適用を試みている。本稿では dPLRM に基づく話者識別法、評価実験について紹介する [2]。

[dPLRM による話者識別]

dPLRM では、混合ガウス分布モデルなどのパラメトリックなモデルを用いずに、カーネル関数を利用して、データのみに基づいて自動的にモデルを選択し、データを高次元空間に写像した上で高性能かつ確率的な判別を行う。

実際、dPLRM では入力データについて、そのクラスを予測する確率ベクトル (クラスごとの予測確率で構成される) が得られる。学習では、学習用の各音声について、その話者を予測する確率ベクトルが現実のものに近くなるように (罰金付

きの対数尤度関数が最大となるように)、dPLRM パラメータを推定しておく。テストでは、入力音声について、各フレームごとに予測確率ベクトルを計算し、その積和が最大となる話者を判定結果とする。

[評価実験]

本方法の性能は、テキスト独立型話者識別実験において評価した。比較のために、GMM による実験も行った。実験に用いたデータは、男性10名が複数の文章や単語を、約9ヶ月に渡る4時期に発声した音声である。サンプリング周波数は16kHzで、特徴ベクトルは26次元 (12次元のメルケプストラム、正規化対数パワー、それらの一次回帰係数) であり、25.6ms のハミング窓をかけて10ms ごとに抽出した。学習では、各話者ごとに3文章 (計約12秒) を用いた。テストでは学習とは異なる時期に発声した15単語を個別に用いた。各文章テキストは各話者共通であるが、学習とテストでは異なる。予備実験から、dPLRM のカーネル関数としては9次の多項式関数を、GMM による実験では24混合ガウス分布モデル (diagonal covariance) を用いた。

平均識別率とその信頼区間は dPLRM は $88.7 \pm 2.9\%$ 、GMM は $84.7 \pm 3.5\%$ であり、dPLRM に基づく本方法は GMM と比べて同等以上の性能を示した。この結果は本方法が少ないデータで各話者を有効に識別できることを示している。

今後は大規模なデータベースを用いて、本方法の評価を行うとともに、dPLRM が生成する各クラスの確率推定量を利用して、話者照合や音声区間の検出を行う方法についても検討していく予定である。

参考文献

- [1] K. Tanabe, "Penalized Logistic Regression Machines: New methods for statistical prediction 1," ISM Cooperative Research Report 143, pp. 163-194, 2001..
- [2] T. Matsui and K. Tanabe, "Probabilistic Speaker Identification with dual Penalized Logistic Regression Machine," Proc. ICSLP, 2004.