Language Recognition via Sparse Coding

Overview
- Language Recognition (LRE)
  - Discriminative task to identify spoken language in speech utterances
  - Acoustic signal processing techniques and learning algorithms crucial in modelling distinguishing characteristics of language
- Approach
  - Extract low-level spectral features of speech as input to sparse coding
  - Improve discriminative quality of sparse-coded speech features via maximum a posteriori (MAP) adaptation for sparse coding dictionary
- Result
  - Outperforms Lincoln l-vector pipeline developed for NIST LRE 2015 on subset comprising Arabic and Chinese clusters

Sparse Coding Background
- Unsupervised method to learn efficient representation of data using small number of basis vectors (or atoms) from dictionary \( D \)
- While solving for sparse representation of input, sparse coding also learns \( D \)
- Two forms
  - \( L_1 \)-regularized LASSO (least absolute shrinkage and selection operator) and LARS (least angle regression) min \( \| x - D y \|_2^2 + \lambda \| y \|_1 \) s.t. \( \| y \|_1 \leq T, \| D \| \)
  - \( L_1 \)-regularized matching pursuit (OMP) min \( \| x - D y \|_2^2 \) s.t. \( \| y \|_1 \leq S \).

Approach
- Low-level feature extraction
  - Compute shifted delta cepstra (SDC) by warping spectral features, passing through the_scal and REEN filters, and normalising
- Vanilla sparse coding (VSC)
  - Classical semi-supervised approach
  - Unsupervised high-level feature learning for sparse coding and dictionary learning
  - Sparse-coded features are pooled and applied for supervised training of classifiers (e.g., linear SVM)
- Adaptive sparse coding (ASC)
  - Improves VSC by adapting VSC dictionary \( D \) to observe-specific dictionary \( D_a \) in unsupervised procedure
  - Using both \( D \) and \( D_a \), compute two sparse codes
  - As-threshold difference vector of the two sparse codes are used to train classifiers

Experiments
- Task: NIST Language Recognition Evaluation (LRE) 2015
  - Aims to examine average performance of language recognition system that can classify target language accurately within a well-defined language clusters for a given speech sample
  - LRE 2015 covers 23 different languages in Arabic, Chinese, English, French, Hindi, and Tibetan clusters
  - Our evaluation focuses on Arabic and Chinese clusters only

Results: Individual Pipelines
- Classical semi-supervised
  - VSC by OMP-1024 sparse coding on SDC 0.2066
  - VSC by LARS-1024 sparse coding on SDC 0.2179
  - Adaptive VSC by OMP-1024 sparse coding on SDC 0.1993
  - Adaptive VSC by LARS-1024 sparse coding on SDC 0.1967

- We evaluated both forms of sparse coding LARS and OMP that train 1024 dictionary atoms
- Adaptive sparse coding with LARS-1024 is our best scheme
- Significantly outperforms Vector-SDC baseline

Youngjune L. Gwon, William M. Campbell, Douglas Sturim, MIT Lincoln Laboratory and H. T. Kung, Harvard University