Discriminative Non-negative Matrix Factorization for Single-Channel Speech Separation

Zi Wang
Mentors: Dingshao Lu, Fei Sha

Introduction
Speech separation (Cocktail-party problem)
- Goal: Segregating each stream of sound from mixed speech of many speakers.
- Application: Robust speech recognition: preprocessing noisy or multi-speaker speech data
  - Improve speech quality: boosting signal noise ratio for targeted speech

Nonnegative matrix factorization
- Intuitions
  - Represent speech signals with nonnegative magnitudes of their mel spectrum
  - Model mixed signal’s spectrum as additive sum of each individual source’s spectrum

- Models
  - Let \( D_{i1}, D_{i2}, \ldots, D_{iK} \) denote speaker i’s speech prototypes (e.g., one for each phoneme’s spectrum), \( S_i \) denotes the input signal’s spectrum of speaker i
  - Minimize the difference between input signals and linear combinations of those prototypes for each speaker
  \[
  F(D, H) = \sum_i K L(S_i \parallel D_i H_i)
  \]

- Learning
  - How to learn prototypes D without knowing H?
    - Iteratively learn D and H for each speaker
    - Update rules:
      \[
      H_i \leftarrow H_i \cdot \frac{d_i^T S_i / d_i H_i}{\sum_i d_i H_i}
      \]
      \[
      D_i \leftarrow D_i \cdot \frac{s_i H_i^T / s_i H_i}{\sum_i s_i H_i}
      \]

Discriminative NMF
- Intuitions
  - Reconstructed speech from clean conditions should also be optimal under other interfering conditions
  - Learning jointly all prototypes and consider the sparsity of H

- Models
  - Let \( S_{ij} \) denotes the mixed signal’s spectrum of speaker i and j
  - Let \( \hat{S}_{ij} \) denotes the reconstruction of the mixed signal’s spectrum
  \[
  F(D, H) = \sum_i K L(S_i \parallel D_i H_i) + \sum_{i,j} K L(S_{ij} \parallel \hat{S}_{ij}) + \lambda \sum_i H_i
  \]
  \[
  S_{ij} = [D_i \quad D_j] \times [H_i \quad H_j]
  \]

- Pairwise speakers
  - Limit the speakers involved during training
  - Easily adapt to multiple speakers

- Optimization algorithm
  - Optimize each speaker’s prototypes alternatively
  \[
  \begin{align*}
  a_i & = a_i - \alpha \left( \sum_{j \neq i} \frac{s_i H_j^T / s_i H_i}{\sum_j s_i H_j} \cdot \frac{d_i^T S_i / d_i H_i}{\sum_i d_i H_i} \right) \\
  a_j & = a_j + \alpha \left( \sum_{i \neq j} \frac{s_i H_i^T / s_i H_i}{\sum_i s_i H_i} \cdot \frac{d_j^T S_j / d_j H_j}{\sum_j d_j H_j} \right) \\
  \end{align*}
  \]

- Evaluation
  - Analyze the prototypes and reconstruction coefficients to gain further insight

Current approaches
- Non-negative matrix factorization (NMF)
  - Model non-negative data using parts-based, additive representations
  - Exploit speaker-specific parts to separate mixed speech
- Sparse Non-Negative Matrix Factorization (SNMF)
  - Extend NMF by sparsely combining parts
  - Estimate over-complete dictionaries
- Limitations
  - Learn parts independently
  - Does not adapt to other speakers’ interference

Our approaches
- Main Idea
  - Discriminative non-negative matrix factorization (DNMF)
    - Learn parts jointly for all speakers
    - Optimize parts to be maximally effective in segregating from other speakers
  - Pairwise DNMF
    - Extend DNMF by distinguishing only pairwise speakers
    - Reduce computational cost

Experiment setup
- The Grid Corpus
  - 34 speakers and 1000 sentences per speaker
  - half of the 1000 sentences for each speaker are used for training and the other half for evaluation
- Evaluation
  - tune parameters and validate on development set(half of the evaluation set)
- Results
  - Outperform NMF in improving SNR
  - DNMF vs. NMF
    - No significant improvement
    - Why similar results?
      - the dictionary D from training and the activity H from testing are different
      - the reconstruction, DH, are similar.
    - Pairwise DNMF vs. DNMF
      - Compare one pair of speaker (different gender)
      - Slightly better than DNMF, need further investigation

Conclusion
- We have developed a new method for speech separation.
  - The key idea is to learn speaker-specific parts discriminatively.
  - Our method yields promising results, improving the popular approach NMF.
  - Our method is applicable to other problems where NMF is used.

Selected References