

Speaker Recognition

- Deep speaker vectors (d-vectors)
- Binary speaker embedding
- Max-margin metric learning
- Short utterance speaker recognition
- Others

- Deep speaker vectors (d-vectors)
 - Text-dependent
 - Segment pooling
 - Dynamic time warping
 - Text-independent
 - Semi text-independent
 - Score combination

- Binary speaker embedding
 - I-vector / LDA-vector \rightarrow b-vector
 - LSH-based binary embedding

$$P[h(x_i) = h(x_j)] = 1 - \frac{1}{\pi} \theta(x_i, x_j)$$

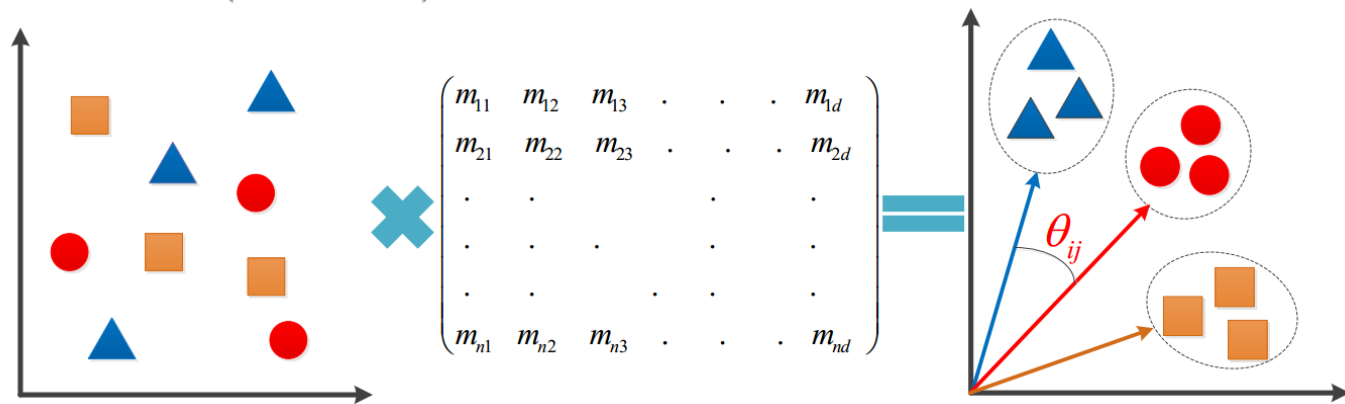
- Variable-sized blocking training

$$l(h, h^+, h^-) = [\|h - h^+\|_H - \|h - h^-\|_H + 1]_+$$

- Verification / Identification

- Max-margin metric learning
 - Metric learning: to learn a projection M.
 - Distance metric: $d(w_1, w_2) = \frac{\langle \mathbf{w}_1, \mathbf{w}_2 \rangle}{\sqrt{\|\mathbf{w}_1\| \|\mathbf{w}_2\|}}$.
 - Goal: to discriminate true speakers and imposters.
 - Max-margin objective function:

$$\mathcal{L}(M) = \sum_{(\mathbf{w}, \mathbf{w}^+, \mathbf{w}^-) \in \mathcal{S}} \max\{0, \delta - d(M\mathbf{w}, M\mathbf{w}^+) + d(M\mathbf{w}, M\mathbf{w}^-)\}$$



- Short utterance speaker recognition
 - Speech unit classes based sub-region modeling
 - Phonetic knowledge clustering
 - Data-driven clustering
 - I-vector + PLDA model
 - DNN + i-vector structure
 - Model synthesis
 - Sub-region UBMs
 - Cohort speakers

- Others

- Research

- Long-term speaker verification
 - Gender-dependent feature extraction
 - Community detection model
 - Discriminative feature selection

- Four patents

感 谢