Max-Margin Metric Learning for Speaker Recognition

Lantian Li

CSLT / RIIT, Tsinghua University

lilt@cslt.riit.tsinghua.edu.cn

Co-work with ***Chao Xing, Dong Wang and Thoms Fang Zheng***

http://arxiv.org/abs/1510.05940

Outline

- Introduction
- Max-Margin Metric Learning
- Experiments
- Conclusions

Introduction

- Speaker recognition
  - ✧ to verify claimed identities of speakers
  - ✧ wide applications
- I-vector model
  - ✧ speaker and session variances → a low-dimensional subspace
  - ✧ discriminative model → WCCN/NAP/LDA/PLDA

Introduction

- PLDA (Probabilistic linear discriminant analysis)

  ✧ Advantages

  ✓ training objective function → discriminative ability

  ✓ Gaussian prior assumption → a generative model

  ✧ Shortcomings

  ✓ Euclidian distance ***vs.*** Cosine distance

  ✓ Binary decision task ***vs.*** Multi-class discrimination training

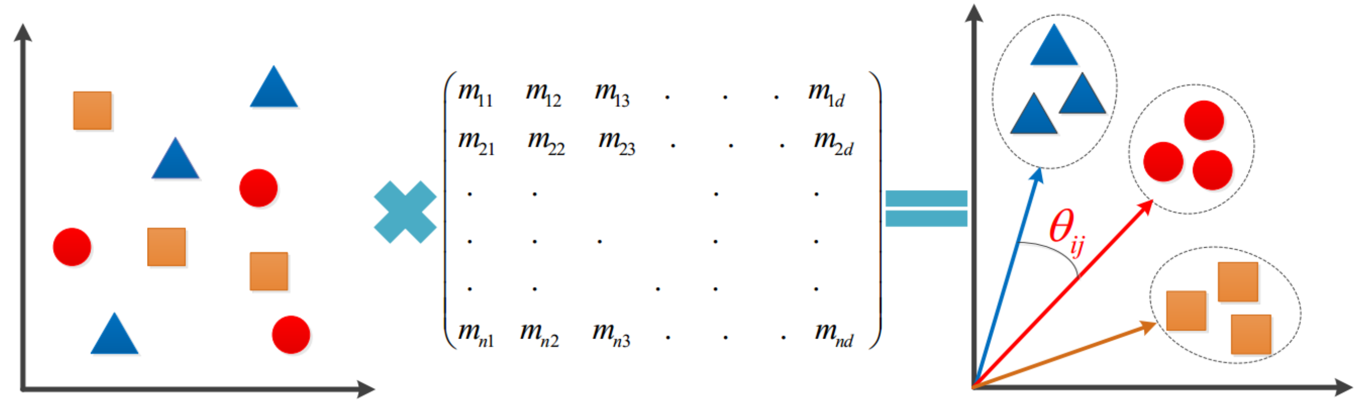
  ✓ Strong Gaussian assumption

Outline

- Introduction
- **Max-Margin Metric Learning**
- Experiments
- Conclusions

Max-margin metric learning

- MMML (Max-margin metric learning)
  - ✧ Metric learning: to learn a projection ***M***.
  - ✧ Distance metric: $d(w_1, w_2) = \dfrac{< \mathbf{w}_1, \mathbf{w}_2 >}{\sqrt{||\mathbf{w}_1||||\mathbf{w}_2||}}.$
  - ✧ Goal: to discriminate true speakers and imposters.



Max-margin metric learning

- MMML (Max-margin metric learning)

  ✧ A contrastive triple (w, w^+, w^-)

  ✧ Max-margin objective function

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

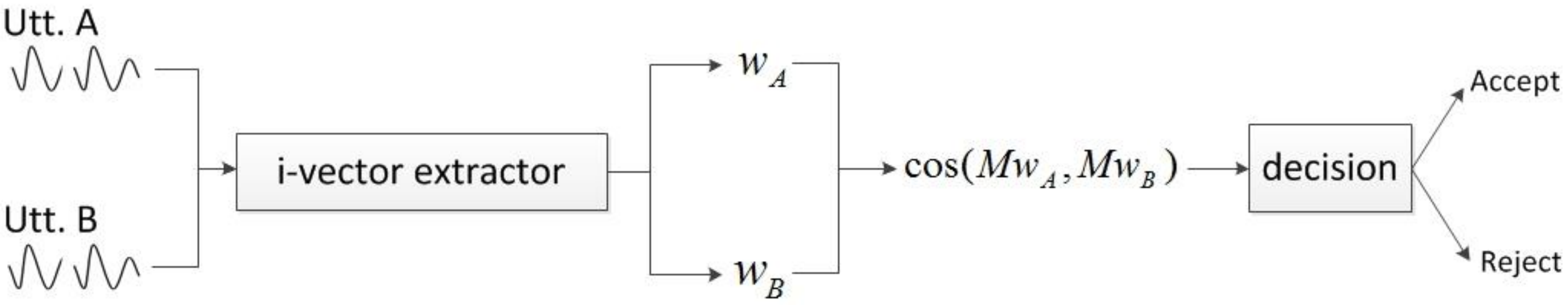
  ✧ SGD algorithem

$$M^t = M^{t-1} + \epsilon \frac{\partial \mathcal{L}}{\partial M}$$

Max-margin metric learning

- MMML (Max-margin metric learning)

  ✧ Test trials



Outline

- Introduction
- Improved Deep Feature Learning
- **Experiments**
- Conclusions

Experiments

- Database

  ✧ Dev. Set: Fisher → 7196 speakers and 13287 utterances.

  ✧ Eva. Set: NIST SRE 08 → short2-short3-trials (59343 trials).

- Experimental Setup

  ✧ 39-dims MFCCs, 2048 Gaussian components.

  ✧ 400-dims i-vector, 150-dims LDA projection space.

  ✧ 400*150-dims MMML projection matrix.

Experiments

- Basic results

Condition	Cosine	LDA	PLDA	MMML
C1	29.34	22.11	18.57	**13.65**
C2	4.78	**1.19**	1.79	**1.19**
C3	29.66	22.65	18.70	**14.12**
C4	18.92	12.91	14.41	**10.66**
C5	20.31	14.42	**10.58**	11.42
C6	12.47	10.75	**9.42**	11.25
C7	7.73	5.58	**4.06**	6.08
C8	7.37	5.52	**4.21**	5.26
Overall	25.58	20.96	19.13	**15.64**

Table 1. *EER results on NIST SRE 2008 core test. The best results are shown in bold face for each condition .*

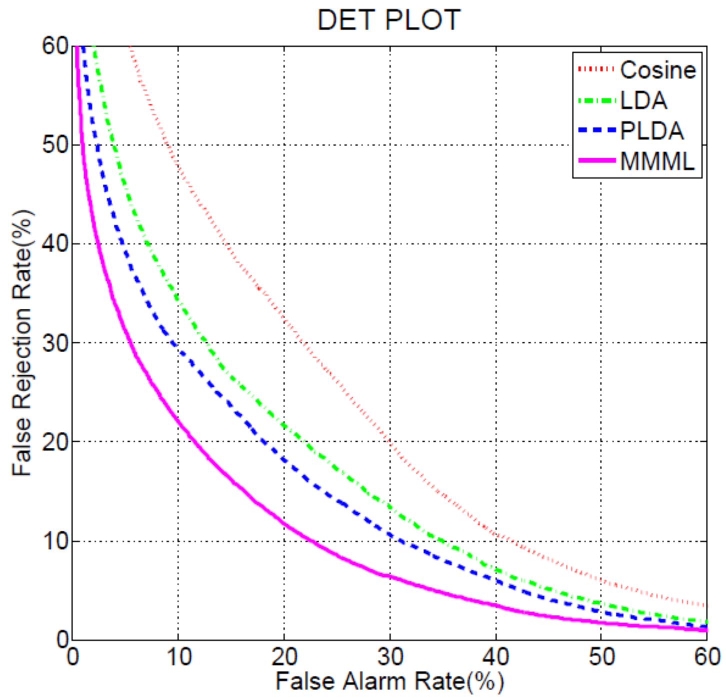


Fig. 2. *The DET curves on the NIST SRE 2008 overall test condition.*

Experiments

- Score fusion

Condition	Cosine	LDA	PLDA	MMML
C1	29.34	22.11	18.57	**13.65**
C2	4.78	**1.19**	1.79	**1.19**
C3	29.66	22.65	18.70	**14.12**
C4	18.92	12.91	14.41	**10.66**
C5	20.31	14.42	**10.58**	11.42
C6	12.47	10.75	**9.42**	11.25
C7	7.73	5.58	**4.06**	6.08
C8	7.37	5.52	**4.21**	5.26
Overall	25.58	20.96	19.13	**15.64**

Table 1. *EER results on NIST SRE 2008 core test. The best results are shown in bold face for each condition .*

Condition	LDA + MMML	PLDA + MMML
C1	16.45	16.22
C2	0.60	0.90
C3	17.04	16.53
C4	10.06	10.96
C5	11.54	9.38
C6	10.31	9.03
C7	5.32	4.06
C8	5.00	3.68
Overall	17.84	17.67

Table 3. *EER results with score fusion, where the interpolation factor for MMML is chosen to be* 0.4.

Outline

- Introduction
- Improved Deep Feature Learning
- Experiments
- **Conclusions**

Conclusions

- Max-margin metric learning
  - ✧ simple linear transform
  - ✧ the training criterion of max-margin
  - ✧ cosine distance metric
- As simple as LDA
- Comparable or even better than PLDA

Thank you

Happy New Year