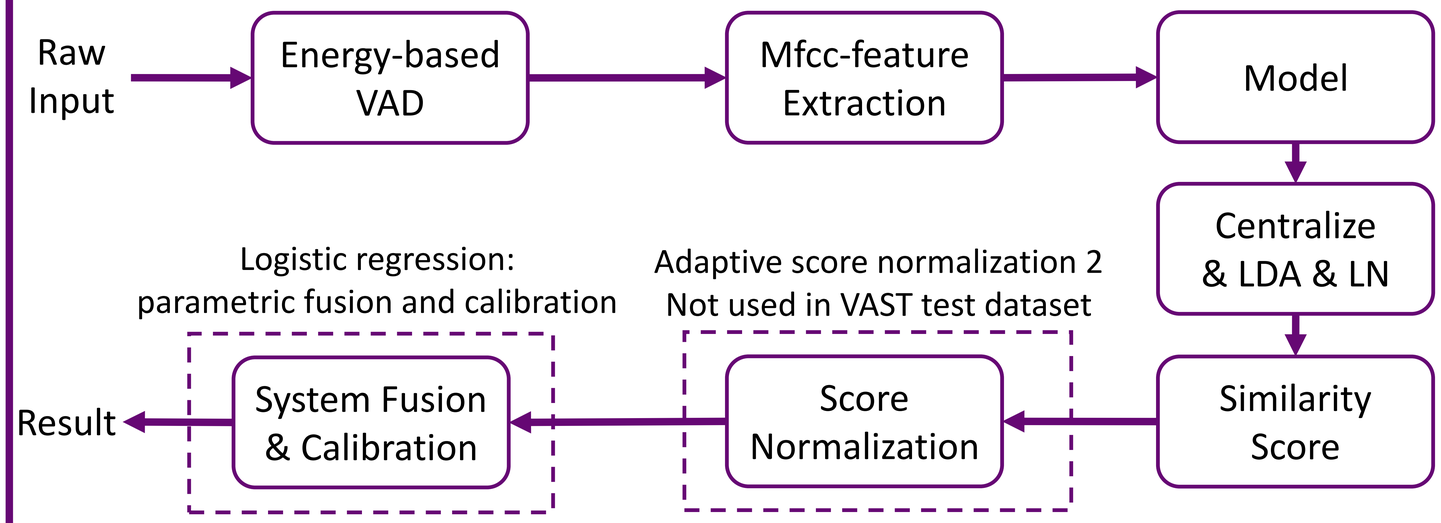
THU-SPMI System For NIST 2018 Speaker Recognition Evaluation

Yutian Li, Zhijian Ou
Speech Processing and Machine Intelligence (SPMI) Lab, Tsinghua University, Beijing, China.
yutian-l16@mails.tsinghua.edu.cn, ozj@tsinghua.edu.cn

Tsinghua University
Department of Electronic Engineering

System Pipeline



Our Contributions

Similarity Score

- Joint Bayesian: $x_{ij} = \mu_i + \epsilon_{ij}$
- Kaldi PLDA: $\overline{x_i} = \mu_i + \overline{\epsilon_i}$

Model

- Model 1: GMM i-vector
- Model 2: DNN i-vector
- Model 3: Kaldi x-vector
- Model 4: Angular x-vector $L = \frac{1}{N}\sum_{n=1}^{N} -\log\left(\frac{e^{\|\mathbf{x}^{(n)}\|\varphi\left(\theta_{y_n}^{(n)}\right)}}{e^{\|\mathbf{x}^{(n)}\|\varphi\left(\theta_{y_n}^{(n)}\right)} + \sum_{j\neq y_n} e^{\|\mathbf{x}^{(n)}\|\cos\left(\theta_j^{(n)}\right)}}\right)$
- Model 5: Cosine x-vector $L = \frac{1}{N}\sum_{n=1}^{N} -\log\left(\frac{e^{s\left(cos\left(\theta_{y_n}^{(n)}\right)-m\right)}}{e^{s\left(cos\left(\theta_{y_n}^{(n)}\right)-m\right)} + \sum_{j\neq y_n} e^{s\left(\cos\left(\theta_j^{(n)}\right)\right)}}\right)$

* Red color indicates our contributions in SRE18. Model 4 and model 5 are implemented by Pytorch.

Min_C of CMN2 development data (after score norm)

	Model 1	Model 2	Model 3	Model 4	Model 5
Joint Bayesian	**0.665**	**0.645**	**0.478**	0.481	**0.480**
Kaldi PLDA	0.683	0.674	0.493	**0.470**	0.496

Result

Fused System

Set	Model	Similarity	condition	EER[%]	min_C	act_C
CMN2	1,2,3,4	Joint Bayesian	dev	7.01	0.401	0.411
			eval	7.02	0.441	0.443
VAST	1,2,3,5 (Whole) 3,5 (Seg)	Kaldi PLDA	dev	5.35	0.370	0.453
			eval	17.14	0.760	0.864

* Whole: use test utterance as a whole in scoring
* Seg: cut test utterance into 2s segments, score each segment, and then use the largest score as the final score

Single System

Set	Model	Similarity	condition	EER[%]	min_C	act_C
CMN2	3	Joint Bayesian	dev	7.57	0.478	0.489
			eval	7.87	0.494	0.496
VAST	2 (Whole)	Kaldi PLDA	dev	11.11	0.379	0.564
			eval	14.92	0.709	0.719

Reference

1. Wang, Yiyan, Haotian Xu, and Zhijian Ou. "Joint bayesian gaussian discriminant analysis for speaker verification." ICASSP, 2017.
2. Li, Yutian, Feng Gao, Zhijian Ou, and Jiasong Sun. "Angular Softmax Loss for End-to-end Speaker Verification." ISCSLP, 2018.
3. Wang, Hao, et al. "CosFace: Large margin cosine loss for deep face recognition." CVPR, 2018.
4. Matejka, Pavel, et al. "Analysis of score normalization in multilingual speaker recognition." Interspeech, 2017.