

Experiments setup:**Data:**

[Out of set genre]: interview, singing, movie, vlog

Training data: CN800 – [Out of set genre]

Test data: [Out of set genre] in CN200

Training strategy:

$$z = F_{\theta}(x)$$

While not convergence:

$$\theta' = \theta - a \nabla_{\theta} F(G_s; \theta)$$

$$\theta'' = \theta - b \nabla_{\theta'} F(G_q; \theta')$$

$$\theta = \theta''$$

<p>x: original vector z: new vector θ: parameter of $F(x)$ G_s: support set genre G_q: query set genre a,b: learning rate</p>

In each iteration, support genre and query genre are randomly selected from training set. Note that support set and query set have same speakers.

F(x) Network:

name	Input x output
input	512
dense1	512 x 512 (embedding)
dense2	512 x 512
dense3	512 x 512
Softmax/Arcsoftmax	512 x 800

Experiments result:

TDNN:

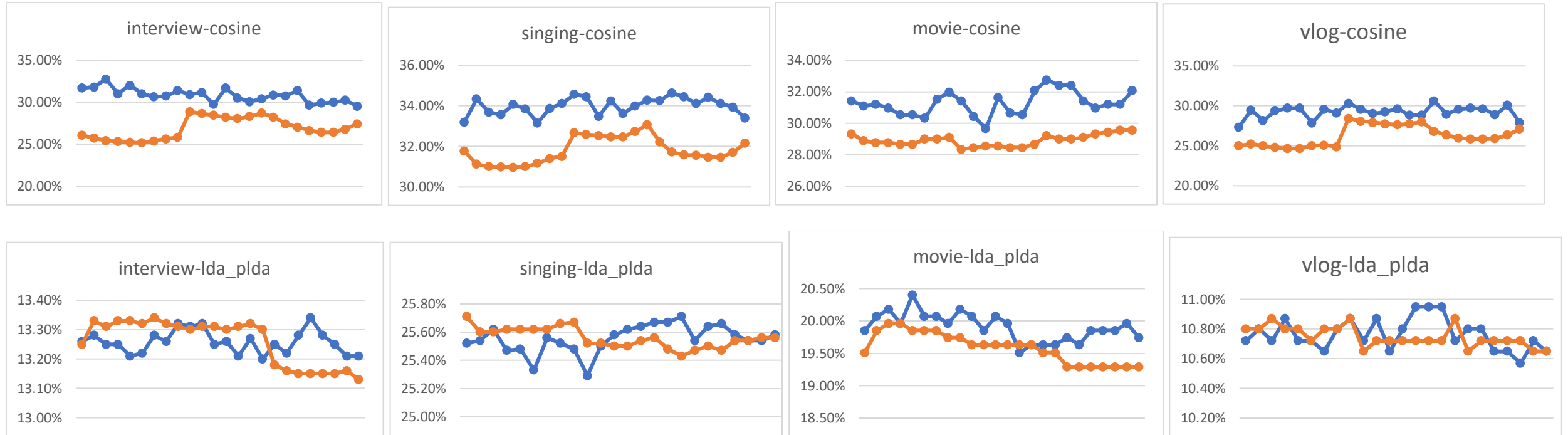
Original vector:

EER	interview	singing	movie	Vlog
cosine	37.17	35.71	33.3	32.17
Lda_plda	13.28	25.54	19.63	10.65

Meta vs DNN:

Blue: DNN

Red: Meta



SODA model(resnet+arcsoftmax):

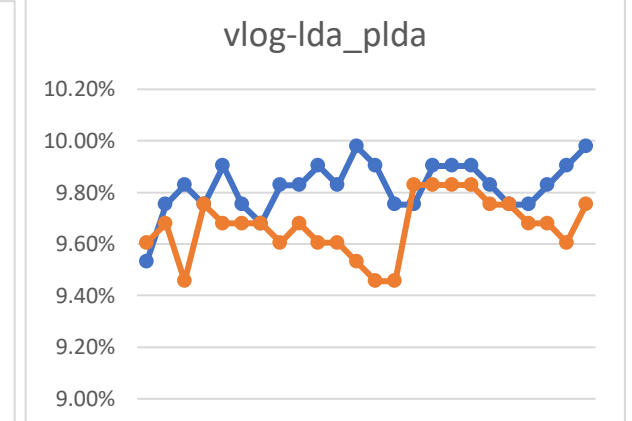
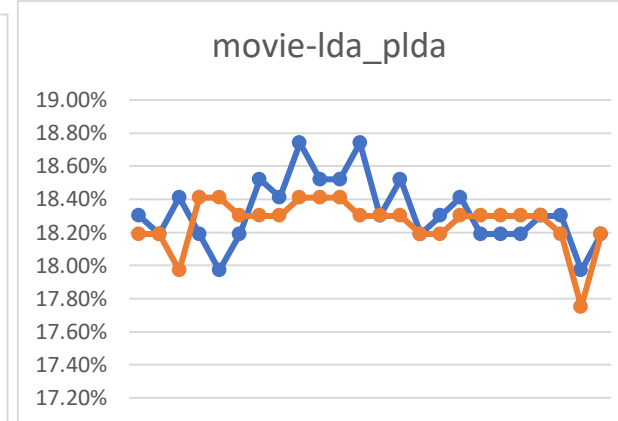
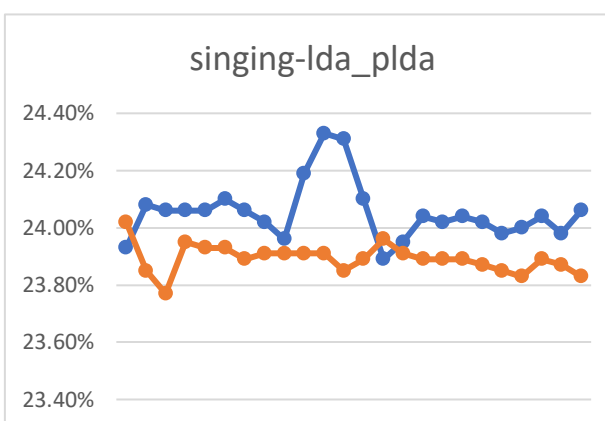
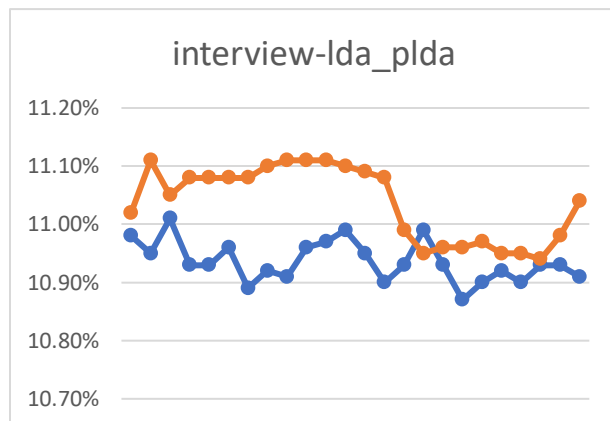
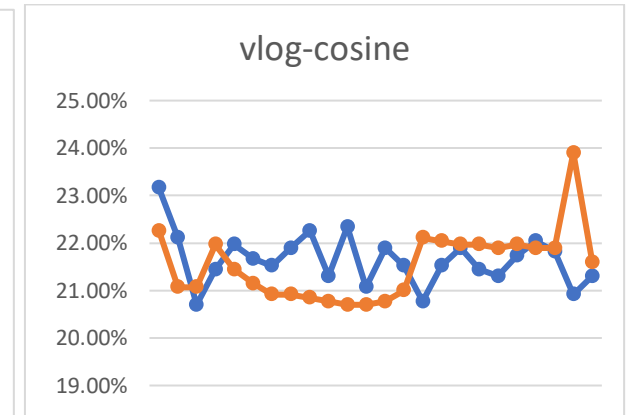
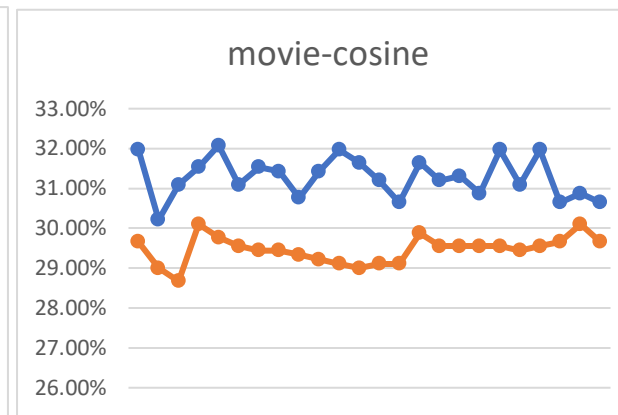
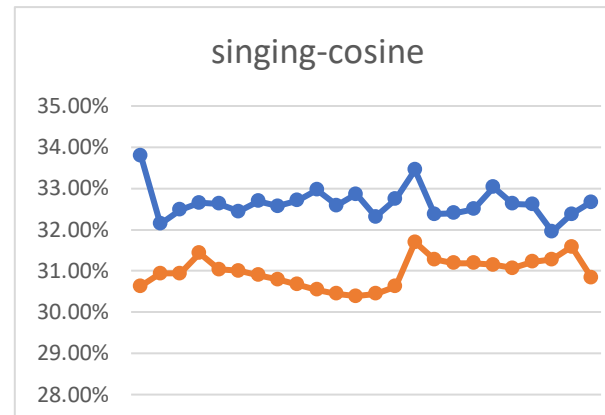
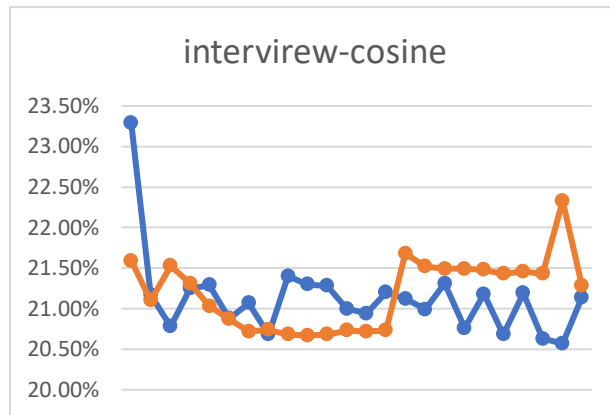
Original vector:

EER	interview	singing	movie	Vlog
cosine	25.68	35.08	33.08	27.70
Lda_plda	11.02	24.12	17.97	9.61

Meta vs DNN:

Blue: DNN

Red: Meta



SODA model twice running:

Original vector:

EER	interview	singing	movie	Vlog
cosine	25.68	35.08	33.08	27.70
Lda_plda	11.02	24.12	17.97	9.61

Meta vs DNN:

Blue: DNN

Red: Meta

