

实验——Mult-scale information

Target: 希望结合不同层的语言信息以获取更好的结果（理由：不同的层代表着不同级别的语言属性）。

基线: x-vector、i-vector（默认 xvector）

训练数据: train_25h

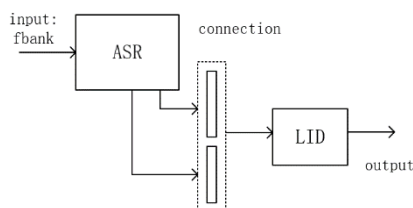


fig.5 网络结构

具体实现: 如 fig. 5 所示，选取不同层的输出，认为其具有不同的语言信息，然后进行拼接输入至 xvector 进行语种识别。实验结果如下：

参数说明:

(A,B) = connection(A, B)

直接将 A 与 B 前后拼接。

(A,B-pca*) = connection(A,B: pca 至*维)

先将 B 层 pca 至*维 再与 A 层连接。

tt: pca 参数 --normalize-mean=true --normalize-variance=true

ff: pca 参数 --normalize-mean=false --normalize-variance=false

f: fbank

x: output-xent.linear

p: prefinal-l

td10: tdnn10l

t8: tdnn8l

实验 1——单独不同层的比较

output layer	in-domain				Out-domain			
	pos	cos	lr	lda-plda	pos	cos	lr	lda-plda
fbank	58.00	58.21	61.70	65.92	27.98	21.84	31.05	36.11
output-xent.linear	96.89	96.72	96.81	96.19	61.79	58.06	64.51	59.75
output.linear	96.40	96.08	96.51	96.11	52.29	51.00	53.87	53.37
prefinal-l	96.91	96.31	96.53	97.01	53.25	53.30	57.85	58.97
tdnn8l	88.93	88.88	89.83	86.74	42.20	35.04	40.60	32.73

结论：（1）output-xent.linear 层的识别率最高。

（2）不同层识别率的不同证明了不同层所包含的语言属性不同。

(x, x-50)	96.43	96.49	96.85	96.87	62.13	59.42	65.11	58.18
(x, x-30)	96.70	96.63	96.42	95.32	60.72	57.95		58.64

结论：黄色部分为比基线好的部分。

(1) 识别率总体相差不大。

(2) 在 (x, *-50) 时，集外识别率比基线高一点点。在后端为 cos 的情况下高 **1.36%**。

实验 5——利用 lda 进行降维

output layer	in-domain				Out-domain			
	pos	cos	lr	lda-plda	pos	cos	lr	lda-plda
fbank	58.00	58.21	61.70	65.92	27.98	21.84	31.05	36.11
output-xent.linear	96.89	96.72	96.81	96.19	61.79	58.06	64.51	59.75
(f, x)	93.19	91.92	91.67	94.28	42.67	41.13	42.73	44.20
(f, x-lda9)	69.95	69.96	70.97	77.15	29.35	24.15	32.41	36.75
(f, x-lda9-2)	68.21	67.97	70.23	73.72	-	-	-	-
(x, p)	96.53	92.86	96.62	94.20	58.17	54.36	57.21	61.41
(x, p-lda9)	96.42	96.43	96.02	95.86	59.46	58.02	59.94	60.66
(x, p-lda9-2)	96.45	96.10	96.07	94.24	60.06	58.01	59.30	59.71
(x, p-lda9, t8-lda9)	96.57	96.27	96.96	96.13	55.24	54.01	56.52	52.88

具体做法：(x, p-lda9)：将 prefinal-1 层向量归一化后降维至 9 维与 output-xent.linear 层拼接。

(x, p-lda9-2)：将 prefinal-1 层降维至 9 维与 output-xent.linear 层拼接（未做向量归一化处理）。

结论：(1) 在做 lda 前归一化比未做归一化的识别率高一点点。

(2) fbank 特征经过 lda 拼接之后的识别率相对于 fbank 的提升了。

(3) output-xent.linear 层的 lda 拼接在集外测试集降低了 4% 左右。

(4) 在集外测试集中，经过 lda 降维之后的识别率相对于未经过降维的 con、con1 识别率分别降低了 7.45%、0.75%。

实验 7 i-vector system

pca 降维之后，输入至 ivector 系统

output layer	in-domain				Out-domain			
	pos	cos	lr	lda-plda	pos	cos	lr	lda-plda
mfcc	-	69.65	71.43	73.67	-	24.26	31.94	29.31
(f, p-pca)								

总结：(1) 由于每一层的识别率不同，所以每一层包含的语言属性也确实不同。

(2) 表格中黄色部分为集外数据集中识别率超过基线的部分，也可以猜测，在某些情况下，实验的 target 可以实现。

存在问题:

- (1) ASR model 的上下文为 24。不同层语言属性应该是有区分度的。
- (2) 是否是训练集时间太短 (总 25h) ?
- (3) 如果需要语言属性之间的互补信息来提高识别率, 那理应 (x, t8) 稍微比 (x, p) 要好 (因为层数间隔多), 但实验结果来看并不理想。是否因为每一层的抽象层度不同, 直接拼接的方式存在问题?
- (4) 分类器问题

Next:

- (1) Continue to explore
在 (x, *-50) 时, 集外识别率比基线高一点点。在后端为 cos 的情况下高 **1.36%**。是否在其它层也存在这种现象 (待测试)。
- (2) score fusion
- (3) Multi-language BN

附件:

[结果有误]实验3——利用 pca 进行降维: 将与 xent.linear 相拼接的 prefinal-1 层进行 pca 降维

3.1 利用 pca 做降维

output layer	in-domain				Out-domain			
	pos	cos	lr	lda-plda	pos	cos	lr	lda-plda
(x, p)	96.53	92.86	96.62	94.20	58.17	54.36	57.21	61.41
(x, p-pca200-tt)	91.86	85.80	94.20	87.32	39.51	53.80	53.80	28.27
(x, p-pca100-tt)	95.69	92.98	95.08	93.31	-	-	-	-
(x, p-pca50-tt)	-	-	-	-	-	-	-	-

结论: 将 prefinal-1 层进行 pca 降维与 xent.linear 拼接后识别率更差, check pca

3.2 Recheck pca, 结合 tsne 图进行分析

(1) 将与 xent 相拼接的 prefinal-1 层进行 pca 降维至 256 (原始维度), 之后在与 xent 层进行拼接

output layer	in-domain				Out-domain			
	pos	cos	lr	lda-plda	pos	cos	lr	lda-plda
(x, p)	96.53	92.86	96.62	94.20	58.17	54.36	57.21	61.41
(x, p-pca256-tt)	88.49	83.50	92.47	73.35	38.39	26.63	53.32	20.67
(x, p-pca256-tf)	59.55	53.20	76.20	51.65	10.53	04.79	47.10	02.95
(x, p-pca256-ft)	88.97	84.43	88.03	77.42	34.82	25.63	57.12	20.85
(x, p-pca256-ff)	58.01	67.06	82.78	41.14	14.05	07.23	44.79	08.45

(2) 将不同层拼接之后进行 pca 降维

output layer	in-domain				Out-domain			
	pos	cos	lr	lda-plda	pos	cos	lr	lda-plda
(f, x)	93.19	91.92	91.67	94.28	42.67	41.13	42.73	44.20
(f, x)-pca49	20.62	22.01	24.36	21.60	-	-	-	-

(3) 单独将 output-xent.linear 层 pca 降维

output layer	in-domain				Out-domain			
	pos	cos	lr	lda-plda	pos	cos	lr	lda-plda

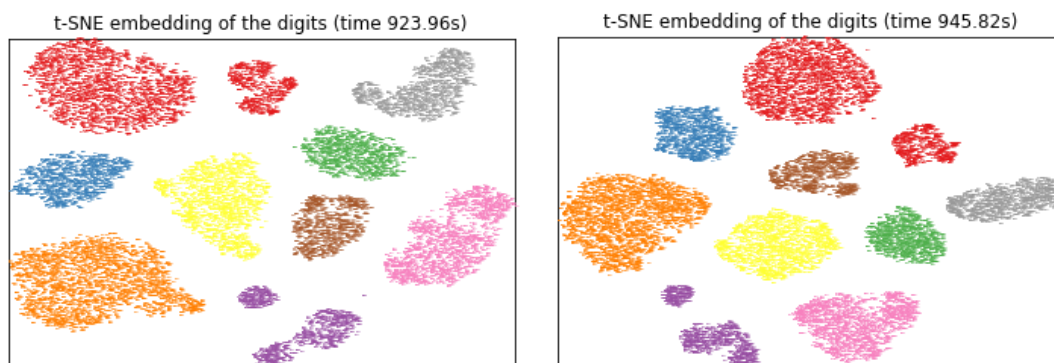
fbank	58.00	58.21	61.70	65.92	27.98	21.84	31.05	36.11
output-xent.linear	96.89	96.72	96.81	96.19	61.79	58.06	64.51	59.75
prefinal-l	96.91	96.31	96.53	97.01	53.25	53.30	57.85	58.97
f-pca40-tt	27.51	30.33	30.53	29.19	-	-	-	-
x-pca256-tt	21.35	18.22	23.40	17.26	-	-	-	-
x-pca256-tf	19.60	16.44	19.83	16.16	-	-	-	-
x-pca256-ft	18.87	18.90	21.60	17.26	-	-	-	-
x-pca256-ff	19.53	10.23	20.93	18.38	-	-	-	-
p-pca256-tt	15.46	16.01	17.89	17.08	-	-	-	-

结论: pca 降维都使识别率降低。tsne 图见表格下, pca 似乎并没有问题, 输入至 xvector 之后区分也很明显。

Fig. 6 output-xent.linear 层进行 pca 前后输入至 xvector 之后的 tsne 图。

Before

After



(4) Recheck-2 pca

output layer	in-domain				Out-domain			
	pos	cos	lr	lda-plda	pos	cos	lr	lda-plda
output-xent.linear	96.89	96.72	96.81	96.19	61.79	58.06	64.51	59.75
x-pca256-tt	96.59	95.87	96.88	94.73				

结论: that's ok.