

d-vector

是一种文本相关识别。一个很大的优点是可以与最新的语音识别系统融合

输出one-hot N维向量(唯一一个非零因子表明说话人身份)

用DNN最后一个隐层的输出代替softmax输出层有几个原因。首先，我们可以通过修剪输出层减小DNN模型尺寸，这也使我们能够使用大量的说话者数据而不增加DNN的大小。其次，我们观察到了更好的泛化性能。

给一个说话人的言论集 $X_S = \{O_{S1}, O_{S2}, \dots, O_{Sn}\}$ ，有观测值 $O_{Si} = \{o1, o2, \dots, om\}$ ，注册过程如下：先用每个言论 O_{Si} 的观测值 o_j 与文本内容一起喂给DNN，获得最后一个隐层的输出后L2正则化。收集言论 O_{Si} 所有观测值的输出后累积作为言论 O_{Si} 的d-vector。最终用 X_S 所有言论的d-vector平均值代表该说话人的特征。

计算得分时也是用余弦距离。

作者把 dvector 与 ivector 模型融合后得到了比只用 ivector 好得多的结果