

PLDA简介 以及PLDA在说话人识别中的应用

张陈昊

2013.11.11

Introduction

- ◎ Linear Discriminant Analysis (LDA)
 - > 寻找类间和类内比例最大的方向，投影到该方向的子空间上，增加区分性
- ◎ Principal Component Analysis (PCA)
 - > 数据的特征数非常多，只有一部分特征是真正我们感兴趣和有意义的，而其他特征或者是噪音或者冗余。
 - > 从所有的数据中找出方差最大的方向进行投影，增加区分性

Fisherface/Fishervoice

◎ LDA / PCA 的区别

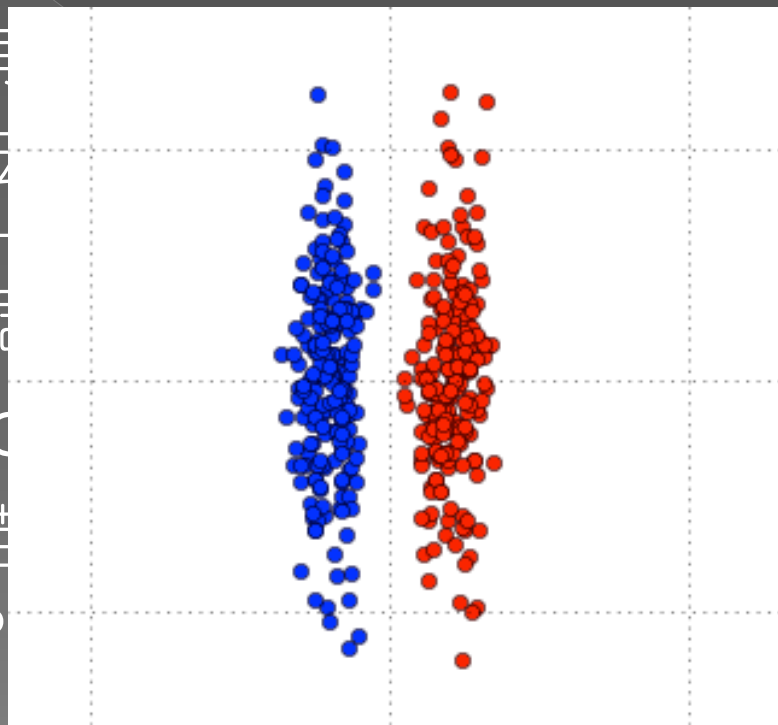
- > LDA是有监督的
- > PCA算法是无监督的
- > PCA算法把原数据映射到一维空间，只考虑了数据分布的方差

先监督的
别), 把原数据
, LDA算法则考

◎ FisherFace

- > 更好地描述了人脸的特征
- > 避免了PCA的缺陷

+ PCA



Probabilistic LDA

◎ PLDA

- > FisherFace对应的概率分布版本
- > 最早出现在图像领域
- > 相比PCA,LDA等降维算法, 有更好的效果

PLDA模型

假设如下

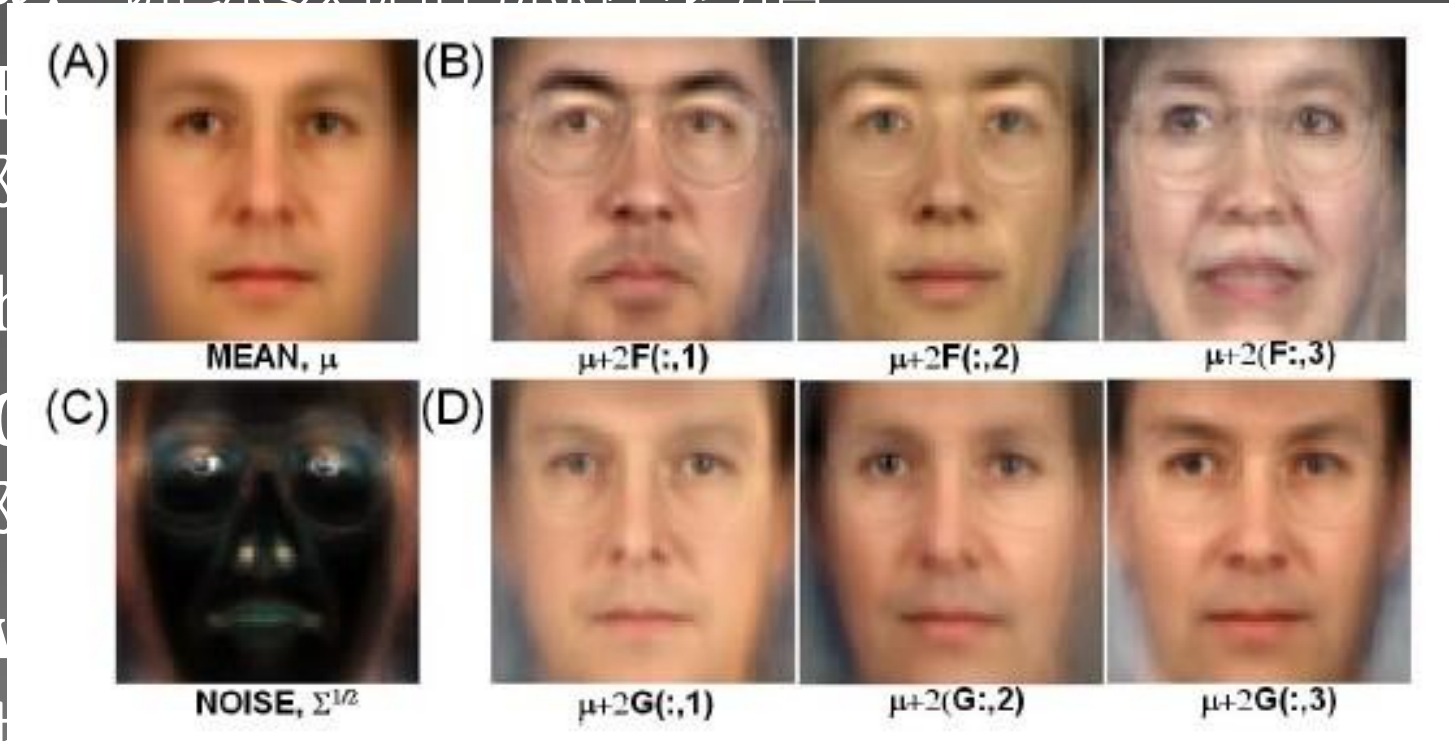
- 训练数据：I个人物，每人J张人脸图像，并把来自第i个人的第j张图像定义为 x_{ij}
- 模型如下：

$$x_{ij} = \mu + \mathbf{F}h_i + \mathbf{G}w_{ij} + \epsilon_{ij}$$

- > 该模型包含两部分： $\mu + \mathbf{F}h_i$ 只与人有关系，而和具体图像没关系； $\mathbf{G}w_{ij} + \epsilon_{ij}$ 不止和人有关系，还和每张图有关系

$$x_{ij} = \mu + \mathbf{F}h_i + \mathbf{G}w_{ij} + \epsilon_{ij}$$

○ μ : 训练数据的总体均值



○ ϵ_{ij} : 残差，分部服从同一对角协方差矩阵 Σ

PLDA

- 训练：
 - > 模型参数： $\theta = \{\mu, F, G, \Sigma\}$
- 测试
 - > 比较 \mathbf{h}_i

$$\begin{bmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \\ \vdots \\ \mathbf{x}_N \end{bmatrix} = \begin{bmatrix} \mu \\ \mu \\ \vdots \\ \mu \end{bmatrix} + \begin{bmatrix} \mathbf{F} & \mathbf{G} & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{F} & \mathbf{0} & \mathbf{G} & \dots & \mathbf{0} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \mathbf{F} & \mathbf{0} & \mathbf{0} & \dots & \mathbf{G} \end{bmatrix} \begin{bmatrix} \mathbf{h} \\ \mathbf{w}_1 \\ \mathbf{w}_2 \\ \vdots \\ \mathbf{w}_N \end{bmatrix} + \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_N \end{bmatrix}$$

Gaussian PLDA

- 公式的基本形式和图像中的PLDA没有区别，讲I个人的J张图像替换成I个人的J段语音
- 说话人识别中的一些假设
 - > 一般假设 x_{ij} ， μ 服从协方差矩阵 Σ （一般定为对角阵）相同的高斯分布
 - > \mathbf{h}_i 和 \mathbf{w}_{ij} 是低维的正态分布
 - > 残差 ϵ_{ij} 满足均值为0，协方差矩阵为 Σ 的正态分布
- P. Kenny在2011年对模型进行了修改
$$x_{ij} = \mu + \mathbf{F}\mathbf{h}_i + \epsilon_{ij}$$

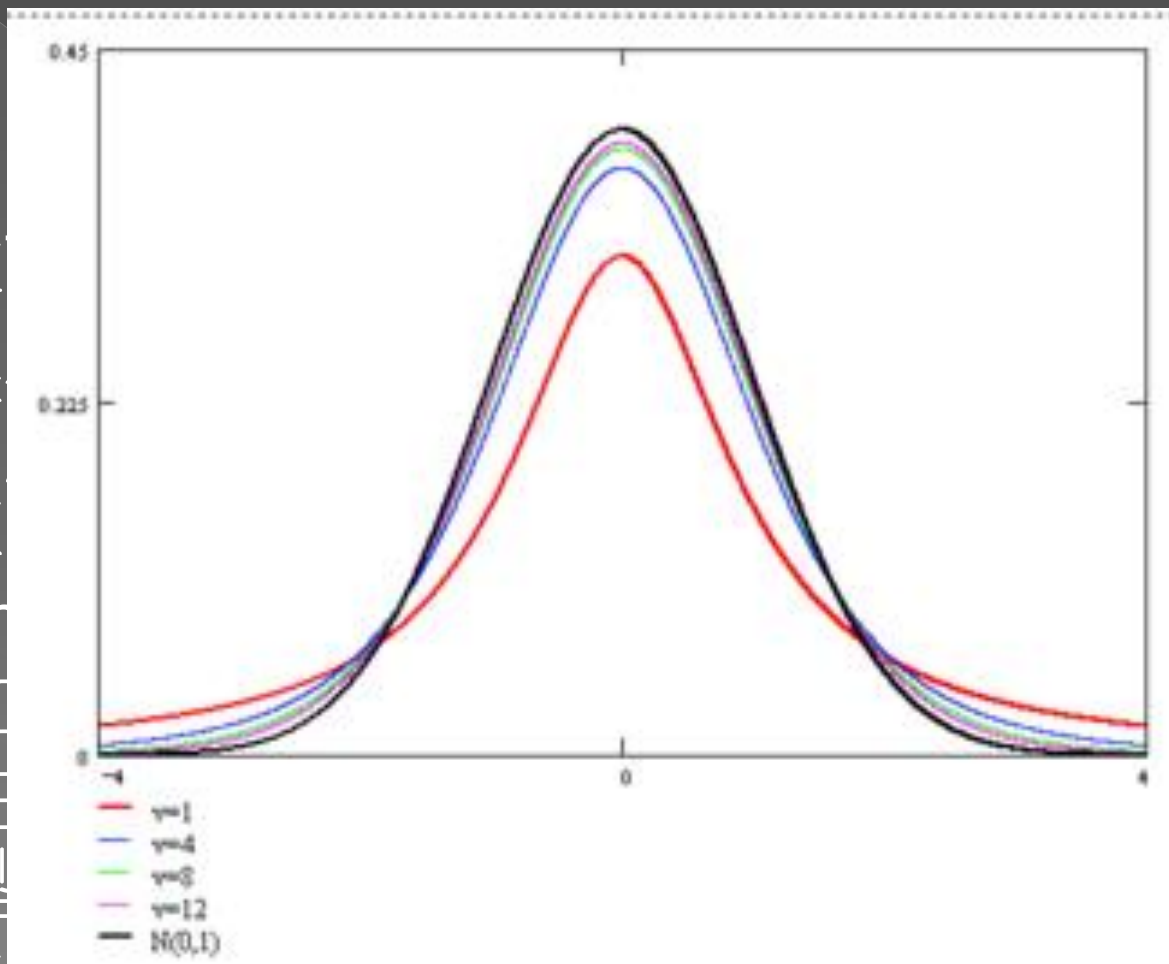
Heavy-Tailed PLDA

- ◎ T分布

- > 标准
- > 如果
- 自由
- 实际效

- ◎ P. Ker
- Σ 在影
- 分布由
- 的模型

- > 将前提假设的高斯分布，替换为T分布



；在
的边

假定的
而T
更好

HT-PLDA和G-PLDA的比较

- ◎ 实验结果说明，通过调节模型，G-PLDA和HT-PLDA可以达到的性能基本相当，HT-PLDA的性能会有一些改进，但是不明显（相对5%）
- ◎ HT-PLDA计算开销更大，但是系统性能较为稳定
 - Although HT-PLDA is able to successfully cope with the development and evaluation mismatch induced by the i-vector extraction procedure. A small improvement is observed after transforming the i-vectors. [4]

参考文献

1. Prince S J D, Elder J H. Probabilistic linear discriminant analysis for inferences about identity[C]//Computer Vision, 2007. ICCV 2007. IEEE 11th International Conference on. IEEE, 2007: 1-8.
2. Garcia-Romero D, Espy-Wilson C Y. Analysis of i-vector Length Normalization in Speaker Recognition Systems[C]//INTERSPEECH. 2011: 249-252.
3. Kenny P. Bayesian speaker verification with heavy tailed priors[C]//Speaker and Language Recognition Workshop (IEEE Odyssey). 2010.
4. Garcia-Romero D, Espy-Wilson C Y. Analysis of i-vector Length Normalization in Speaker Recognition Systems[C]//INTERSPEECH. 2011: 249-252.
5. Senoussaoui M, Kenny P, Brümmer N, et al. Mixture of PLDA Models in i-vector Space for Gender-Independent Speaker Recognition[C]//INTERSPEECH. 2011: 25-28.