Image Retrieval
BOW, VLAD, FV

字节跳动人工智能实验室　杨成
GitHub: https://github.com/ironyoung

BOW

- Bag of Words. Based on clustering（聚类）
- 聚类方法汇总
  - Partitioning: 建立数据分割，相同标准评价聚类结果
  - Model-based: 假设分布模型，寻找最优分布
  - Reduction: 先降维，再聚类

BOW - Partitioning

- eg. kmeans

- Kmeans 目的：把 N 个点（可以是样本的一次观察或一个实例）划分到 k 个聚类中，使每个点都属于离他最近的聚类中心

$$\underset{\mathbf{S}}{\arg\min} \sum_{k=1}^{K} \sum_{\mathbf{x} \in S_k} \| \mathbf{x} - \boldsymbol{\mu}_k \|^2$$

- Kmeans 算法步骤：

    - 分配 (Assignment)：将每个观测点分配到聚类中，使组内平方和最小。所以把观测点分配到离它最近得均值点即可。尽管理论上可能分配到多个聚类

    - 更新(Update)：对上一步得到的每个聚类，以聚类中观测值图心作为新的均值点

BOW - Partitioning

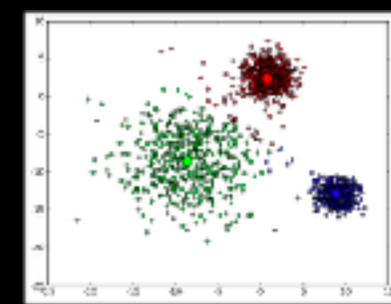
- Kmeans in image retrieval：
  - 提取 local descriptor (SIFT, SURF …)
  - descriptor kmeans
  - image -> descriptors -> centroids = words

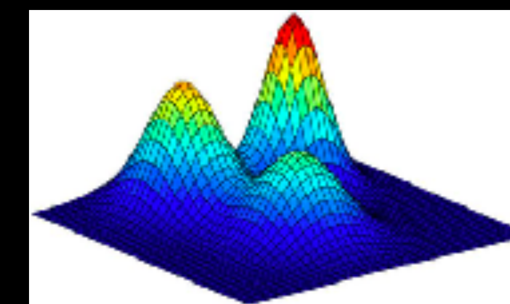
BOW - Partitioning

- Kmeans Tricks：
  - 多中心聚类（权重分配）
  - 多次随机初始点
  - 分段聚类

BOW - Model-based

- eg. GMM (Gaussian mixture model)





- k 个高斯模型混合在一起，每个点出现的概率是若干个高斯模型混合的结果

- 多元正态分布的概率密度函数

$$f(x;\mu,\sigma)=\frac{1}{\sigma\sqrt{2\pi}}\exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$$

标准差　平均值

$$f_{\mathbf{x}}(\mathbf{x})=\frac{1}{\sqrt{(2\pi)^d|\boldsymbol{\Sigma}|}}\exp\left(-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})\right)$$

多元维度　协方差矩阵　均值向量

高斯模型个数，又称 Component 个数

- 记 $\mathcal{N}(\mathbf{x}|\boldsymbol{\mu},\boldsymbol{\Sigma})=f_{\mathbf{x}}(\mathbf{x})$，则 GMM 概率密度函数为 $p(\mathbf{x}|\mu_k,\Sigma_k)=\sum_{k=1}^{K}w_k\mathcal{N}(\mathbf{x}|\boldsymbol{\mu}_k,\boldsymbol{\Sigma}_k)$

高斯分布权重 $\sum_{k=1}^{K}w_k=1$

BOW - Model-based

- 假定数据由 GMM 生成出来，则根据数据推导出 GMM 的概率分布（已知结果，反推先验模型），GMM 的 K 个 Component 对应 K 个 cluster

- 已知：概率密度函数形式（GMM, $p(\mathbf{x})$）；“推导” 参数的过程：参数估计

- 哪些参数需要估计？$\boldsymbol{\theta} = \boldsymbol{w}, \boldsymbol{\mu}, \boldsymbol{\Sigma}$

- 参数估计的方法？最大化似然函数

- 什么是似然函数？对 N 个观察（ $\mathbf{x}_1 \ldots \mathbf{x}_n$ ）而言，似然函数

$$L(\theta \mid \mathbf{x}_1 \ldots \mathbf{x}_n) = p(\mathbf{x}_1 \ldots \mathbf{x}_n \mid \theta) = \prod_{n=1}^{N} p(\mathbf{x_n} \mid \theta)$$

BOW - Model-based

- 最大化似然函数 $L(\theta \mid \mathbf{x}_1 \dots \mathbf{x}_n) = p(\mathbf{x}_1 \dots \mathbf{x}_n \mid \theta) = \prod_{n=1}^{N} p(\mathbf{x_n} \mid \theta)$

- 浮点数下溢，所以使用对数形式

$$\log L(\theta \mid \mathbf{x}_1 \dots \mathbf{x}_n) = \log \left(\prod_{n=1}^{N} p(\mathbf{x_n} \mid \theta) \right) = \sum_{n=1}^{N} \log \left(\sum_{k=1}^{K} w_k \mathcal{N}(\mathbf{x}_n \mid \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k) \right)$$

- 求解参数方法
  - 最大化 $\arg\max L(\theta \mid \mathbf{x}_1 \dots \mathbf{x}_n)$： 求导 / 偏导求极值
  - 多参数 $\boldsymbol{\theta} = \boldsymbol{w}, \mu, \Sigma$： EM 算法

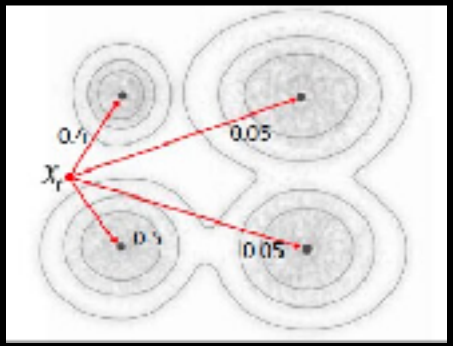
BOW - Model-based

- 似然函数 $\log L(\theta \mid \mathbf{x}_1 \ldots \mathbf{x}_n) = \sum_{n=1}^{N} \log \left(\sum_{k=1}^{K} w_k \mathcal{N}(\mathbf{x}_n \mid \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k) \right)$

- 似然函数对 μ_k 求偏导，并令其极值 = 0，可得

$$-\sum_{n=1}^{N} \boxed{\frac{w_k \mathcal{N}\left(\mathbf{x}_n \mid \mu_k, \Sigma_k\right)}{\sum_i^K w_i \mathcal{N}\left(\mathbf{x}_n \mid \mu_i, \Sigma_i\right)}} \Sigma_k \left(\mathbf{x}_n - \mu_k\right) = 0$$

$= \gamma(n, k) \Rightarrow$ clustering



- 记 $N_k = \sum_{n=1}^{N} \gamma(n, k)$，上式化简后得 $\mu_k := \frac{1}{N_k} \sum_{n=1}^{N} \gamma(n, k) \mathbf{x}_n$

- 同理，对 Σ_k, w_k 分别求偏导数，化简得到

- $\Sigma_k := \frac{1}{N_k} \sum_{n=1}^{N} \gamma(n, k)(\mathbf{x}_n - \mu_k)(\mathbf{x}_n - \mu_k)^T$ $\qquad w_k := \frac{N_k}{N}$

BOW - Model-based

- EM 算法步骤 $$\arg\max \sum_{n=1}^{N} \log\left(\sum_{k=1}^{K} w_k \mathcal{N}(\mathbf{x_n} \mid \boldsymbol{\mu_k}, \boldsymbol{\Sigma_k})\right)$$
- Initialization（Kmeans 初始化 GMM）：$\theta = \boldsymbol{w}, \boldsymbol{\mu}, \boldsymbol{\Sigma}$
- E step： $\boldsymbol{w}, \boldsymbol{\mu}, \boldsymbol{\Sigma} \Rightarrow \boldsymbol{\gamma}$
- M step： $\boldsymbol{\gamma} \Rightarrow \boldsymbol{w}^{new}, \boldsymbol{\mu}^{new}$

  $\boldsymbol{\gamma}, \boldsymbol{\mu}^{new} \Rightarrow \boldsymbol{\Sigma}^{new}$
- Evaluation: evaluate log likelihood, may return to E step

BOW - Reduction

- 首先 dimensionality reduction：PCA, random projection, …
- 继续 clustering

VLAD

- 不定长特征样本（eg. SIFT）-> 定长特征
-

$$\boldsymbol{v}_k = \sum_{\boxed{\mathbf{x} \in S_k}} (\mathbf{x} - \boldsymbol{c}_k)$$

第 k 类聚类中心

可以归属于多个类，权重分配

Fisher Vector

- FV 本质是用似然函数的梯度向量表示一副图像

$$\log L(\theta \mid \mathbf{x}_1 \ldots \mathbf{x}_n) = \log \left(\prod_{n=1}^{N} p(\mathbf{x_n} \mid \theta) \right) = \sum_{n=1}^{N} \log p(\mathbf{x_n} \mid \theta)$$

- 各求梯度，FV 维度 (1+2D)*K - 1 $\sum_{k=1}^{K} w_k = 1$ ，各个分量为：

$$\frac{\partial L(\mathbf{x} \mid \theta)}{\partial w_k} = \sum_{n=1}^{N} \left[\frac{\gamma(k,n)}{w_k} - \frac{\gamma(k,1)}{w_1} \right]$$

$$\frac{\partial L(\mathbf{x} \mid \theta)}{\partial \mu_k^d} = \sum_{n=1}^{N} \gamma(k,n) \left[\frac{x_k^d - \mu_k^d}{(\sigma_k^d)^2} \right]$$

- $$\frac{\partial L(\mathbf{x} \mid \theta)}{\partial \sigma_k^d} = \sum_{n=1}^{N} \gamma(k,n) \left[\frac{(x_k^d - \mu_k^d)^2}{(\sigma_k^d)^3} - \frac{1}{\sigma_k^d} \right]$$

Fisher Vector

- 生成式模型关注类条件概率的建模，学 $p(X,Y) \xrightarrow{\text{Bayes}} p(Y|X)$
- 判别式模型则直接关注问题的本身，直接学习 $p(Y|X)$
- FV 是一个生成模型，但是得到 FV 之后接入的分类器是判别式模型，依赖于归一化后的向量
- 如何对于 FV 归一化呢?

Fisher Vector

- Fisher Vector 归一化（概率空间）$G_\theta^{\mathbf{x}} = \frac{\partial \log p(\mathbf{x} \mid \theta)}{\partial \theta}$
- FIM (Fisher Information Vector) $F_\theta = E_{\mathbf{x}}\left[(G_\theta^{\mathbf{x}})^2\right]$
- 归一化后梯度向量 $G_\theta^{\mathbf{x}} = F_\theta^{-1/2} G_\theta^{\mathbf{x}}$
- FIM 一般求近似解

Fisher Vector

- FIM 归一化后的梯度向量

$$f_{w_i} = K(\frac{1}{w_i} + \frac{1}{w_1})$$

$$f_{\mu_i^d} = K\frac{w_i}{(\sigma_i^d)^2}$$

$$f_{\sigma_i^d} = K\frac{2w_i}{(\sigma_i^d)^2}$$

$$\mathcal{G}_{w_k} = f_{w_i}^{-1/2}\frac{\partial L(\mathbf{x}\,|\,\theta)}{\partial w_k}$$

$$\mathcal{G}_{\mu_k^d} = f_{\mu_i^d}^{-1/2}\frac{\partial L(\mathbf{x}\,|\,\theta)}{\partial \mu_k^d}$$

$$\mathcal{G}_{\sigma_k^d} = f_{\sigma_i^d}^{-1/2}\frac{\partial L(\mathbf{x}\,|\,\theta)}{\partial \sigma_k^d}$$

- concatenate 组成 Fisher Vector $\mathcal{G}$，再进行 $\mathcal{L}2$ 正则化

Reference

- 聚类算法分类：https://blog.csdn.net/abcjennifer/article/details/8170687
- 最大似然、EM 算法
  - https://blog.csdn.net/zouxy09/article/details/8537620
  - https://vividfree.github.io/docs/2016-08-19-introduction-about-EM-algorithm-doc1.pdf
- GMM: https://blog.csdn.net/abcjennifer/article/details/8198352
- VLAD: Aggregating local descriptors into a compact image representation
- Fisher Vector
  - Fisher Information：https://www.zhihu.com/question/26561604
  - Image Classification with the Fisher Vector: Theory and Practice
  - Fisher Kernel: Fisher Kernels on Visual Vocabularies for Image Categorization