Kernel Canonical Correlation Analysis

bxf_hit@163.com

2017 年 3 月 23 日

摘要

本文是对近几日所读的几篇关于 KCCA 的文献所做的总结，算是之前 CCA 一文的续篇。

1 KCCA 原理

之前 CCA 一文中已经说到，KCCA 的提出是为了解决 CCA 的非线性相关问题；

KCCA 的主要思想就是利用核函数将原来的低维向量升维到高维空间，这样原来在低维的非线性相关属性到高维就会变成线性相关；这点的思想与 SVM 的核函数 f:$R^n \to R^N (n < N)$ 相似，使用核函数的目的就是将低维向量升到高维空间，使原来的线性不可分问题变成线性可分问题。我们将这里用到的核函数抽象为：

$\Phi : X = (X_1 X_2 \cdots X_n) \to \Phi(X) = (\phi_1(X), \phi_2(X) \cdots \phi_N(X))$ 其中 n<N (1)

2 KCCA 求解

类似于 CCA, 我们的目的也是构造高维空间投影向量 $c, d \in R^N$, 使 $X' = c^T X, Y' = d^T Y$ 间的相关系数最大。然而这样直接去求有如下问题：

- 核函数如高斯径向基核函数会将向量升到无穷维，此时 c,d 自然也是无穷的，自然无法求解

- 引入核函数的一个目的就是利用核函数的核矩阵 K, 也就是期望出现 X^TX 型，替换成核矩阵 K，而上面的做法显然没有 X^TX 型，因此很难利用核矩阵，求解比较复杂

重新定义：$P = \begin{bmatrix} \Phi(X1)^T \\ \Phi(X2)^T \\ \vdots \\ \Phi(Xm)^T \end{bmatrix}$ 则 $P^T = \begin{bmatrix} \Phi(X1) & \Phi(X2) & \cdots & \Phi(Xm) \end{bmatrix}$,

同理 $Q = \begin{bmatrix} \Phi(Y1)^T \\ \Phi(Y2)^T \\ \vdots \\ \Phi(Ym)^T \end{bmatrix}$(这里 X,Y 的样本数量是相同的)

令 c = P^{T}a, d = Q^{T}b, 这里 a 相当于通过控制每个高维向量的权重控制 c.（这里用到了 KPCA 的相关证明，空间中的任一向量（哪怕是基向量），都可以由该空间中的所有样本线性表示）

与之对应的：

$$X' = c^TP, Y' = d^TQ \tag{2}$$

类似于 CCA, 我们要最大化的还是：

$$\rho(X', Y') = \frac{cov(X', Y')}{\sqrt{Var(X')}\sqrt{Var(Y')}} \tag{3}$$

其中：

$$Var(X') = c^Tvar(P)c = a^TPvar(P)P^Ta \tag{4}$$

$$Var(Y') = d^Tvar(Q)d = b^TQvar(Q)Q^Tb \tag{5}$$

$$Cov(X', Y') = c^Tcov(P, Q)d = a^TPcov(P, Q)Q^Tb \tag{6}$$

进行下一步前,先对 P,Q 进行归零化处理, 即 $\sum$xi=0,好处就是 $cov(P, Q) = \frac{1}{N-1}P^TQ$,（这个待证明）。

引入核矩阵 K,K$_p$ = PP^T,K$_q$ =QQ^T

于是上面的公式也可以写成：

$$Var(X') = a^T Pvar(P)P^T a = a^T PP^T PP^T a = a^T K_p^2 a \tag{7}$$

$$Var(Y') = b^T Qvar(Q)Q^T b = b^T QQ^T QQ^T b = b^T K_q^2 b \tag{8}$$

$$Cov(X', Y') = a^T Pcov(P, Q)Q^T b = a^T PP^T QQ^T b = a^T K_p K_q b \tag{9}$$

因此

$$\rho(X', Y') = \frac{cov(X', Y')}{\sqrt{Var(X')}\sqrt{Var(Y')}} = \frac{a^T K_p K_q b}{\sqrt{a^T K_p^2 a}\sqrt{b^T K_q^2 b}} \tag{10}$$

与 CCA 类似，此时的优化问题变成：

Maximize: $a^T K_p K_q b$

Subject to: $a^T K_p^2 a = 1, b^T K_q^2 b = 1$

构造 Lagrangian 等式：

$$L = a^T K_p K_q b - \frac{\lambda}{2}(a^T K_p^2 a - 1) - \frac{\theta}{2}(b^T K_q^2 b - 1) \tag{11}$$

类似于 CCA 的求解过程，最后也可以得到一个求矩阵特征值的方程：

$$K_p K_p a - \lambda^2 K_p K_p a = 0 \tag{12}$$

即：

$$Ia = \lambda^2 a \tag{13}$$

得 $\lambda_1 = \lambda_2 = \cdots = \lambda_k = 1$, 发现所有的 λ 都是相等的，显然模型是过拟合的:)，需要正则化处理。

3 正则化的 KCCA

有两种常见的正则化的方法，一种是对 Lagrange 加入正则项，一种是对限制条件做一下修改。

第一种是引入正则项的 Lagrange 函数为：

$$L = a^T K_p K_q b - \frac{\lambda}{2}(a^T K_p^2 a - 1) - \frac{\theta}{2}(b^T K_q^2 b - 1) + \frac{\eta}{2}(\| a \|^2 + \| b \|^2) \tag{14}$$

同样令 Lagrange 函数的导数为 0，得：

$$K_pK_qb = (\lambda K_p^2 + \eta I)a \tag{15}$$

$$K_qK_pa = (\theta K_q^2 + \eta I)b \tag{16}$$

最后化简可以求得：

$$(K_p + \eta I)^{-1}K_q(K_q + \eta I)^{-1}K_qa = \lambda^2 a \tag{17}$$

又回到了求特征值和特征向量的问题了，已经解决。

第二种是对限制条件作修改，修改后的模型为：

Maximize: $a^TK_pK_qb$

Subject to: $\begin{array}{c}(1-\sigma)a^TK_p^2a + \sigma a^TK_pa = 1 \\ (1-\sigma)b^TK_q^2b + \sigma b^TK_qb = 1\end{array}$

继续构造 Lagrange 函数，最后也可以化简为与方案 1 类似的结果，这里就不详细推导了。

4 小结

KCCA 通过使用核函数升维的思想是原来低维空间内非线性相关的两组属性到高维空间内能够相性相关，从而捕获了一些非线性相关的信息，相对于 CCA 是个很不错的提升。KCCA 的实现还有一些细节没有说明，如怎样 Centering 高维空间的数据，矩阵的分解等，有时间会继续补充。

KCCA 的提出有一定程度是借鉴了 KPCA 的做法，因此时间允许的条件下我会考虑去阅读下 KPCA 相关的论文。

5 参考文献

1. Canonical Correlation Analysis: An Overview with Application to Learning Methods. David R. Hardoon Sandor Szedmak

2. A kernel method for canonical correlation analysis. Shotaro Akaho

3. Kernel Independent Component Analysis. Francis R. Bach,Michael I. Jordan

4. Kernel Principal Component Analysis. Bernhard Scholkopf , Alexander Smola

5. http://www.cnblogs.com/boostable/p/lec_kernel_canonical_correlation_analysis.html

6. http://www.cnblogs.com/jerrylead/archive/2011/06/20/2085491.html