Canonical Correlation Analysis 典型关联分析

bxf_hit@163.com

2017 年 3 月 23 日

摘要

最近由于毕设的需求，对 CCA 模型进行了比较系统的考察，在这里总结成一篇文章供以后查看。

下面开始正文。

1 CCA 原理分析

典型关联分析 (CCA) 是利用综合变量对之间的相关关系来反映两个多维随机变量之间的整体相关性的多元统计分析方法。

举个简单的例子, 我们想考察一 个人解题能力 X(解题速度 x1, 解题正确率 x2) 与他/她的阅读能力 Y(阅读速 度 y1, 理解程度 y2) 之间的关系; 正常的思路自然是利用 pearson 相关系数，然而这里 X,Y 是多维随机变量，相关系数衡量的是两个单独的随机变量的相关程度，因此无法直接衡量，让我们看看 CCA 是怎么做的。

CCA 的做法就是对于多维随机变量 $X = (X_1X_2....X_m), Y = (Y_1Y_2....Y_n)$ 寻找两个投影向量 a,b, 分别与 X,Y 相乘得到

$$X' = a^TX, Y' = b^TY \tag{1}$$

然后使用 X’ 与 Y’ 的 Pearson 相关系数

$$\rho(X', Y') = \frac{cov(X', Y')}{\sqrt{Var(X')}\sqrt{Var(Y')}} = \frac{E[(X' - \mu_x)(Y' - \mu_y)]}{\sqrt{Var(X')}\sqrt{Var(Y')}} \tag{2}$$

来度量 X 和 Y 的关系, 我们期望寻求一组最优的解 a 和 b 使得相关系数 $\rho(X', Y')$ 最大, 这里的 μ_x, μ_y 分别是 X ‘和 Y’ 的均值。

随机变量 $X' = a^T X$和$Y' = b^T Y$ 是第一对典型变量。然后寻求一个依然最大化相关但与第一对典型变量不相关的向量；这样就得到了第二对典型变量, 这个步骤会进行 min{m,n}次，m,n 分别是 X,Y 的维度。

因此 CCA 的主要工作就是求 a 和 b 两个投影向量，求得了 a,b, 也就求得了典型变量。

2 CCA 求解过程

给定两组向量 X 和 Y,X 维度为 m, Y 维度为 n, 默认 m <= n。形式化表示如下：

$$T = \begin{bmatrix} X \\ Y \end{bmatrix}, E(T) = \begin{bmatrix} \mu_x \\ \mu_y \end{bmatrix}, \Sigma = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix} \tag{3}$$

其中 Σ 是 T 的协方差矩阵，Σ_{11} 是 X 自己的协方差矩阵，也就是 X 的方差矩阵,Σ_{22} 是 Y 的方差矩阵，Σ_{12} 是 X 与 Y 的协方差矩阵，Σ_{21} 是 Σ_{12} 的转置。

由公式（1）和（3）可以得到：

$$Var(X') = a^T \Sigma_{11} a, Var(Y') = b^T \Sigma_{22} b, Cov(X', Y') = a^T \Sigma_{12} b \tag{4}$$

简要推导一个：

$$Var(X') = Var(a^T X) = \frac{1}{N} \sum (a^T X_i - a^T \mu_x)^2 = a^T \frac{1}{N} \sum (X_i - \mu_x)^2 a = a^T \Sigma_{11} a$$

由公式（2）（4）可得：

$$\rho(X', Y') = \frac{cov(X', Y')}{\sqrt{Var(X')}\sqrt{Var(Y')}} = \frac{a^T \Sigma_{12} b}{\sqrt{a^T \Sigma_{11} a}\sqrt{b^T \Sigma_{22} b}} \tag{5}$$

CCA 的目的就是最大化相关系数 ρ, 容易观察到上面的分式中 a = ka,b = mb 时结果不变, 所以此问题可以演变成下面的有条件约束的最优化问题:

Maximize: $a^T \Sigma_{12} b$

Subject to: $a^T \Sigma_{11} a = 1, b^T \Sigma_{22} b = 1$

构造 Lagrangian 等式:

$$L = a^T \Sigma_{12} b - \frac{\lambda}{2}(a^T \Sigma_{11} a - 1) - \frac{\theta}{2}(b^T \Sigma_{22} b - 1) \tag{6}$$

求偏导，令导数得 0：

$$\Sigma_{12}b - \lambda\Sigma_{11}a = 0 \tag{7}$$

$$\Sigma_{21}a - \theta\Sigma_{22}b = 0 \tag{8}$$

令（7）左乘 a^T,(8) 左乘 b^T, 再有 $a^T\Sigma_{11}a = 1, b^T\Sigma_{22}b = 1$ 得

$$\lambda = \theta = a^T\Sigma_{12}b \tag{9}$$

也就是说 λ 就是 corr(X',Y')，只要求出最大的 λ 即可。

由公式（8）导出 b，再带入（7）中，结合（9）$\lambda = \theta$ 可得:

$$\Sigma_{11}^{-1}\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}a = \lambda^2 a \tag{10}$$

这样只要求出矩阵 $\Sigma_{11}^{-1}\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}$ 的最大的特征值和对应的特征向量就能求得 λ和 a 了，b 也可以求得。

至此我们终于完成了第一对典型变量对应的投影变量的求解。

下面我们进行第 2 对和之后的投影变量的求解:

与之前给出的约束条件有所不同的是，第二对投影变量还要保证:

$$a_2^T\Sigma_{11}a_1 = 0, b_2^T\Sigma_{22}b_1 = 0 \tag{11}$$

也就是 cov(a_1^TX,a_2^TX)=0,cov(b_1^TY,b_2^TY)=0, 即第二对典型变量与第一对是线性无关的。

由此，已知前 k-1 个投影向量，第 k 个投影向量可以表示为如下优化问题:

Maximize: $a_k^T\Sigma_{12}b_k$

Subject to: $a_k^T\Sigma_{11}a_k = 1, b_k^T\Sigma_{22}b_k = 1; a_k^T\Sigma_{11}a_i = 0, b_k^T\Sigma_{22}b_i = 0$ (i=0,1,2$\cdots k-1$)

我们先来求解第二组投影向量 a_2, b_2, 构造 Lagrangian 等式:

$$L = a_2^T\Sigma_{12}b_2 - \frac{\lambda}{2}(a_2^T\Sigma_{11}a_2 - 1) - \frac{\theta}{2}(b_2^T\Sigma_{22}b_2 - 1) + \gamma(a_2^T\Sigma_{11}a_1) + \beta(b_2^T\Sigma_{22}b_1)$$
(12)

求偏导，令偏导为 0 得:

$$\frac{\partial L}{\partial a_2} = \Sigma_{12} b_2 - \lambda \Sigma_{11} a_2 + \gamma \Sigma_{11} a_1 = 0 \tag{13}$$

$$\frac{\partial L}{\partial b_2} = \Sigma_{21} a_2 - \theta \Sigma_{22} b_2 + \beta \Sigma_{22} b_1 = 0 \tag{14}$$

$$\frac{\partial L}{\partial a_1} = \gamma \Sigma_{11} a_2 = 0 \tag{15}$$

$$\frac{\partial L}{\partial b_1} = \beta \Sigma_{22} b_1 = 0 \tag{16}$$

a_2^T 左乘 (13)，b_2^T 左乘 (14)，得:

$$a_2^T \Sigma_{12} b_2 - \lambda a_2^T \Sigma_{11} a_2 + \gamma a_2^T \Sigma_{11} a_1 = 0 \tag{17}$$

$$b_2^T \Sigma_{21} a_2 - \theta b_2^T \Sigma_{22} b_2 + \beta b_2^T \Sigma_{22} b_1 = 0 \tag{18}$$

a_2^T 左乘 (15)，b_2^T 左乘 (16)，得:

$$\gamma a_2^T \Sigma_{11} a_2 = 0 \tag{19}$$

$$\beta b_2^T \Sigma_{22} b_1 = 0 \tag{20}$$

将公式（19）（20）带入（17）（18）中，得:

$$a_2^T \Sigma_{12} b_2 - \lambda a_2^T \Sigma_{11} a_2 = 0 \tag{21}$$

$$b_2^T \Sigma_{21} a_2 - \theta b_2^T \Sigma_{22} b_2 = 0 \tag{22}$$

两边消去 a_2^T，b_2^T 后，得:

$$\Sigma_{12} b_2 - \lambda \Sigma_{11} a_2 = 0 \tag{23}$$

$$\Sigma_{21} a_2 - \theta \Sigma_{22} b_2 = 0 \tag{24}$$

不难发现公式（23）（24）与公式（7）（8）拥有相同的形式，由此同上可推得:

$$\Sigma_{11}^{-1}\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}a_2 = \lambda^2 a_2 \tag{25}$$

因此可以得知，a_2 也是矩阵 $\Sigma_{11}^{-1}\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}$ 的一个特征向量，由于 $a_k^T\Sigma_{11}a_i = 0$ 的限制，a_2和 a_1 显然不可能相同，因此 a_2 就是第二大特征值对应的特征向量，以此类推，可以得知共有 $t = rank(\Sigma_{11}^{-1}\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21})$ 组特征向量，也就是 t 组投影向量，对应特征值 λ_i^2(i=1,2,3$\cdots t$),a_i, b_i 也就是第 i 大相关系数对应的投影向量。CCA 的一个比较神奇的性质就是矩阵 $\Sigma_{11}^{-1}\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}a_2$ 的 t 组特征向量刚好就是要求的投影变量，而且投影后的不同组的典型变量是线性无关的！不得不让人感叹，数学真是个充满艺术的地方:)。

至此，CCA 的求解过程正式结束。

3 CCA 的 matlab 代码实现

了解了 CCA 的原理和求解过程后，CCA 的实现其实就很简单了，只要求出矩阵 $\Sigma_{11}^{-1}\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}$ 的特征值就行了，最后对求出的特征向量就是我们要找的投影向量。下面放上 CCA 的一个官方实现：

`https://github.com/muyeby/GraduationProject/tree/master/matlab_code/CCA`

里面基本的函数已经加了注释，实现也很简洁，可以直接调用。

4 小结

CCA 通过降维让本身难以衡量相关性的两组多维随机变量降到 1 维，再利用 Pearson 系数去衡量两组变量的相关性是一个很有创意的想法，提出后也得到了很多应用；当然 CCA 本身也存在局限性，由于 Pearson 系数也叫线性相关系数，因此对于非线性相关的元素是无法捕捉的，因此后人又提出了 KCCA 等解决方案，下一回会系统介绍下 KCCA.

5 参考文献

1. Canonical correlation analysis: An overview with application to learning methods. David R. Hardoon , Sandor Szedmak and John Shawe-

Taylor

2. http://www.cnblogs.com/boostable/p/lec_canonical_correlation_analysis.html
3. http://www.cnblogs.com/jerrylead/archive/2011/06/20/2085491.html
4. http://www.stat.cmu.edu/~ryantibs/datamining/lectures/10-cor1.pdf
5. Improving Vector Space Word Representations Using Multilingual Correlation. Manaal Faruqui and Chris Dyer