A KERNEL METHOD FOR CANONICAL CORRELATION ANALYSIS

S. Akaho

NATIONAL INSTITUTE OF ADVANCED INDUSTRIAL SCIENCE AND TECHNOLOGY, JAPAN
s.akaho@aist.go.jp

1. Canonical correlation analysis (CCA) is a technique to extract common features from a pair of multivariate data. In complex situations, however, it does not extract useful features because of its linearity. On the other hand, kernel method used in support vector machine (Vapnik, 1998) is an efficient approach to improve such a linear method. In this study, we investigate the effectiveness of applying kernel method to the canonical correlation analysis.
2. Suppose there are pairs of multivariates x and y. We want to find transformations of x and y which extract features commonly embeded in x and y. CCA can find linear transformations which maximize the correlation coefficients between extracted features. However, if the embedding of common features is nonlinear or the relation between features is not Gaussian, CCA sometimes fails to find appropriate features.
3. In this presentation, we propose the kernel canonical correlation analysis (KCCA). Basic idea is as follows: First, the original pair of multivariate variables x and y are mapped into some high demensional Hilbert spaces H_x and H_y respectively, i.e., $\phi_x(x)$ and $\phi_y(y)$. After that, we find linear transformations of $\phi_x(x)$ and $\phi_y(y)$, which can find more complex transformation than CCA.
4. However, there are two problems to be solved, especially because of the dimensionality d of the Hilbert spaces. One is that it is an ill-posed problem if d is larger than sample size. The other is that time complexity to calculate $\phi_x(x)$ and $\phi_y(y)$ may be large. Let a and b be linear coefficients for $\phi_x(x)$ and $\phi_y(y)$ here.
5. In order to solve the former problem, we introduce a quadratic regularization term $\eta(a^2+b^2)/2$ into the cost function of CCA. By the quadratic regularization, it follows that a can be written by a weighted sum of $\phi_x(x_i)$ where x_i is the i-th sample, and b can be written by a weighted sum of $\phi_y(y_i)$. Therefore, $a \cdot \phi_x(x) = \sum_i \alpha_i \phi_x(x_i) \cdot \phi_x(x)$. This fact enables us to use a "kernel trick": Let $k(z, w)$ be a kernel function which is symmetric and positive definite, then there exists $\phi_z(z)$ and $k(z, w) = \phi_z(z) \cdot \phi_z(w)$. Using a kernel, we can calculate $\phi_x(x_i) \cdot \phi_x(x)$ directly without knowing ϕ. This means the complexity problem of calculation is solved as well, because we do not need to calculate ϕ anymore.
6. Consequently, we obtain KCCA: 1. Calculate matrices of kernels $K_x = (k_x(x_i, x_j))$ and $K_y = (k_y(y_i, y_j))$, where k_x and k_y are kernels. 2. Solve the generalized eigen problem, $M\beta = \lambda L\alpha$ and $M^T\alpha = \lambda N\beta$, where $M = (1/n)K_x^T J K_y$, $L = (1/n)K_x^T J K_x + (\eta/\lambda)K_x$, $N = (1/n)K_y^T J K_y + (\eta/\lambda)K_y$ and $J = I - \mathbf{1}^T\mathbf{1}$, $\mathbf{1} = (1, \ldots, 1)^T$.
7. Some simulation results show KCCA recovers nonlinear relation between multivariate variables effectively. However, the performance is sensitive to the regularization constant η. If η is too small, the solution is unstable and there are many features achieves large correlation coefficient, which makes difficult to choose relevant features. On the other hand, if η is too large, all the correlation coefficients tend to be small. Currently, we choose η by cross validation test.
8. We have proposed the extension of CCA by introducing the kernel method. Canonical correlation analysis includes Fisher discriminant analysis (FDA) as a special case. Mika et al. (1999) has proposed the kernel method for FDA, which is also included in our study.

Key words: Canonical correlation analysis, Kernel method, Support Vector Machine

References

Anderson, T.W. (1984) *An Introduction to Multivariate Statistical Analysis — Second edition*, John Wiley & Sons.
Mika, S., Rätsch, G., Weston, J., Schölkpf, B., Müller, K. (1999) Fisher discriminant analysis with kernels, In *Hu, Y.-H. et al.(eds.) Neural Networks for Signal Processing IX*, (pp. 41-48). IEEE.
Schölkopf, B., Smola, A. and Müller, K. (1998) Kernel principal component analysis, In B. Schölkopf et al. (eds), *Advances in Kernel Methods, Support Vector Learning*, MIT Press.
Ramsey, J.O. and Silverman, B.W. (1997) *Functional Data Analysis*, Springer-Verlag.
Vapnik, V.N. (1998) *Statistical Learning Theory*, John Wiley & Sons.