

Data Fusion

Kernel Based Data Fusion and its Application
to Protein Function Prediction in Yeast

Lanckriet, Deng, Cristianini, Jordan, Noble

P. Agius – Spring 2008

Goals

- Combine multiple data types
- Use kernels to combine data types
- Combine kernels by SDP
- Predict yeast protein functions using SVM on 5 types of data

P. Agius – Spring 2008

Previous research

- Treat each data type independently
- Combine multiple datasets using intersections and unions of the overlapping sets of predictions
- Find relationships between data (eg. Sequence similarity for protein sequences, gene co-expression), represent these graphically, then use graph algorithms to extract clusters of data that share similarities in the graphs
- Statistical methods – Eg. Joint likelihood models
- NEW in this paper - Discriminative data fusion. Combine kernel matrices to form a weighted linear combination of kernels

P. Agius – Spring 2008

Methods

Given a set of kernels K_1, \dots, K_m ,
form the linear combination

$$K = \sum_{i=1}^m \mu_i K_i.$$

How can we find the optimal weights μ ?

P. Agius – Spring 2008

Recall SVM

Primal

$$\begin{aligned} \min_{\mathbf{w}, b, \xi} \quad & \mathbf{w}^T \mathbf{w} + C \sum_{i=1}^n \xi_i \\ \text{subject to} \quad & y_i(\mathbf{w}^T \Phi(\mathbf{x}_i) + b) \geq 1 - \xi_i, \quad i = 1, \dots, n \\ & \xi_i \geq 0, \quad i = 1, \dots, n \end{aligned}$$

Other dual formulation

$$\begin{aligned} \max_{\alpha} \quad & \sum_i \alpha_i - \frac{1}{2} \sum_{i,j} \alpha_i \alpha_j y_i y_j x_i x_j \\ \text{s.t.} \quad & \sum_i \alpha_i y_i = 0 \\ & 0 \leq \alpha_i \leq C \end{aligned}$$

Dual

$$\begin{aligned} \max_{\alpha} \quad & 2\alpha^T \mathbf{e} - \alpha^T \text{diag}(\mathbf{y}) K \text{diag}(\mathbf{y}) \alpha \\ \text{subject to} \quad & C \geq \alpha \geq 0, \quad \alpha^T \mathbf{y} = 0, \\ \text{where} \quad & \mathbf{w} = \sum_{i=1}^n \alpha_i \Phi(\mathbf{x}_i) \end{aligned}$$

P. Agius – Spring 2008

Semidefinite Programming

- Convex optimization
- Optimizes a linear objective function in the cone of psd matrices
- A generalization of linear programming
- Scalar linear inequality constraints replaced by linear matrix inequalities (LMIs)
(Eg. $F(\mathbf{x}) \succeq 0$ indicates matrix F must be psd)
- Interior point algorithms used to solve SDP problems

P. Agius – Spring 2008

SDP formulation for kernel combo

$$\begin{aligned}
 & \min_{\mu_i, t, \lambda, \nu, \delta} && t \\
 \text{subject to} &&& \text{trace} \left(\sum_{i=1}^m \mu_i K_i \right) = c, \\
 &&& \sum_{i=1}^m \mu_i K_i \succeq 0, \\
 &&& \begin{pmatrix} \text{diag}(\mathbf{y}) \left(\sum_{i=1}^m \mu_i K_i \right) \text{diag}(\mathbf{y}) & \mathbf{e} + \nu - \delta + \lambda \mathbf{y} \\ \left(\mathbf{e} + \nu - \delta + \lambda \mathbf{y} \right)^T & t - 2C\delta^T \mathbf{e} \end{pmatrix} \succeq 0, \\
 &&& \nu, \delta \geq 0,
 \end{aligned}$$

must be psd

P. Agius – Spring 2008

SDP → QCQP

Authors constrain the kernel μ_i weights to be non-negative.

Each kernel K_i is already psd.

Therefore the linear combo of kernels $K = \sum_{i=1}^m \mu_i K_i$ is automatically psd.

So, ... the SDP can be rewritten as a

quadratically constrained quadratic program (QCQP) as follows:

$$\begin{aligned}
 & \max_{\alpha, t} && 2\alpha^T \mathbf{e} - ct \\
 \text{subject to} &&& t \geq \frac{1}{n} \alpha^T \text{diag}(\mathbf{y}) K_i \text{diag}(\mathbf{y}) \alpha, \quad i = 1, \dots, m \\
 &&& \alpha^T \mathbf{y} = 0, \\
 &&& C \geq \alpha \geq 0.
 \end{aligned}$$

P. Agius – Spring 2008

Results

- 6 kernels
 - Pfam: Protein domain structure
 - Phys: Protein-protein interactions
 - Gen: Genetic interactions
 - TAP: Co-participation in protein complex
 - Exp: Cell cycle gene expression measurements (Pearson correlation)
 - SW: Using Smith-Waterman pairwise sequence comparison algorithm to yeast protein sequences
- kernel method used
 - Gaussian kernel with sigma=0.5

P. Agius – Spring 2008

ROC areas

Kernel	Binary data		Enriched kernels	
	Weight	ROC	Weight	ROC
K_{Pfam}	2.21	.9331	1.58	.9461
K_{Gen}	0.18	.6093	0.21	.6093
K_{Phys}	0.94	.6655	1.01	.6655
K_{TAP}	0.74	.6499	0.49	.6499
K_{Exp}	0.93	.5457	—	.7126
K_{SW}	—	—	1.72	.9180
all	—	.9674	—	.9733

Table 2: **Kernel weights and ROC scores for the transport facilitation class.** The table shows, for both experiments, the mean weight associated with each kernel, as well as the ROC score resulting from learning the classification using only that kernel. The final row lists the mean ROC score using all kernels.

P. Agius – Spring 2008

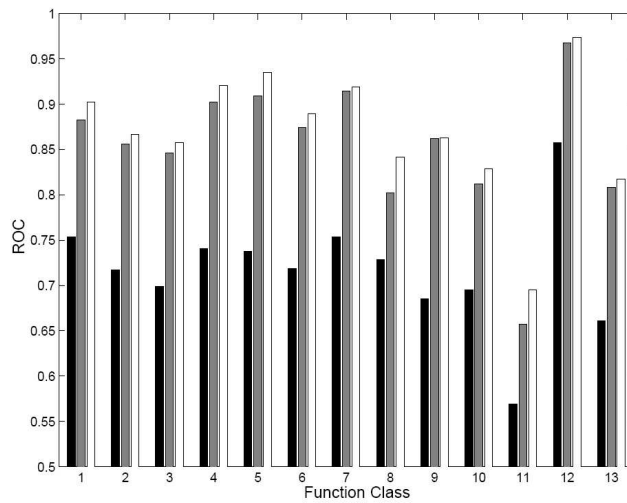


Figure 1: **Classification performance for the 13 functional classes.** The height of each bar is proportional to the ROC score. The standard deviation across the 15 experiments is usually 0.01 or smaller (see supplement), so most of the depicted differences are significant. Black bars correspond to the MRF method of Deng *et al.*; gray bars correspond to the SDP/SVM method using five kernels computed on binary data, and white bars correspond to the SDP/SVM using the enriched Pfam kernel and replacing the expression kernel with the SW kernel.

Deng's MRF is a Markov random field model

Concluding remarks

- Incorporation of different types of biological knowledge
- Different types of data → same kernel format
- Combination of different data using standard kernel format
- Exploitation of semidefinite programming to ensure kernel is psd
- Over to Lempis ...
Data fusion for gene prioritization