

Metric learning pairwise kernel for graph inference with SVM

Jean-Philippe Vert¹ Jian Qiu² William Stafford Noble³

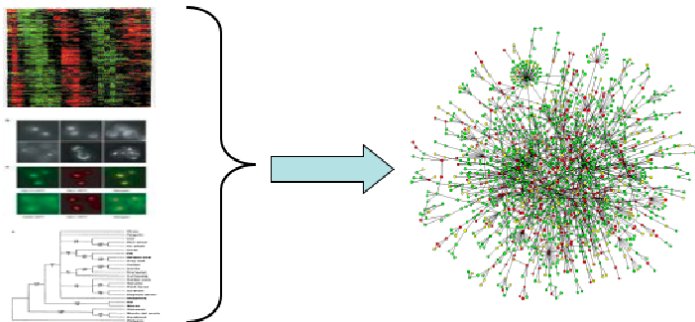
¹Center for Computational Biology
Ecole des Mines de Paris

²Department of Genome Sciences
University of Washington

²Department of Genome Sciences
Department of Computer Science and Engineering
University of Washington

NIPS workshop “Open problems in computational biology”, Whistler,
Canada, December 8th, 2006

Motivation



Data

- Gene expression,
- Gene sequence,
- Protein localization, ...

Graph

- Protein-protein interactions,
- Metabolic pathways,
- Signaling pathways, ...

Unsupervised approaches

The graph is completely unknown

- **model-based** approaches : Bayes nets, dynamical systems,...
- **similarity-based** : connect similar nodes

Supervised approaches

Part of the graph is known

- **Undirect approach** (ML): a pair (a, b) is likely to be connected if it is similar to a pair (c, d) which is connected.
- **Direct approach**: a is likely to be connected to b if they are similar to each other, for a measure of similarity to be optimized

Unsupervised approaches

The graph is completely unknown

- **model-based** approaches : Bayes nets, dynamical systems,...
- **similarity-based** : connect similar nodes

Supervised approaches

Part of the graph is known

- **Undirect approach** (ML): a pair (a, b) is likely to be connected if it is similar to a pair (c, d) which is connected.
- **Direct approach**: a is likely to be connected to b if they are similar to each other, for a measure of similarity to be optimized

Example: supervised undirect approach with SVM

TPPK kernel

- Suppose a kernel K exists for **individual** genes.
- Construct a **tensor product pairwise kernel (TPPK)** between pairs (Ben-Hur and Noble, 2005):

$$K_{TPPK}((x_1, x_2), (x_3, x_4)) = K(x_1, x_3)K(x_2, x_4) + K(x_1, x_4)K(x_2, x_3).$$

- Given a training set of connected and non-connected pairs, train a binary SVM to predict if a new pair is connected or not.

Example: supervised direct approach

Metric learning

- Suppose a kernel K exists for **individual** genes.
- Given a training set of connected and non-connected pairs, find a **new distance metric** such that **connected genes are closer to each other than non-connected pairs**.
- For a new pair, predict an edge is the distance is below a threshold (Vert and Yamanishi, 2005).

- A **new distance metric learning algorithm** for supervised graph inference
- A **relaxation** equivalent to a SVM with a particular kernel for pairs, which we call **metric learning pairwise kernel (MLPK)**.
- Therefore the possibility to:
 - use **out-of-the-box SVM implementation** for supervised direct graph inference
 - easily **combine direct and undirect approaches** by kernel combination.

Objective function

- Suppose genes are described as vectors $x \in \mathbb{R}^d$
- After a linear mapping $\Phi(x) = Ax$ the square Euclidean distance is:

$$\begin{aligned}d_M(x, x') &= (x - x')^\top M (x - x') \\ &= \text{tr} \left(M (x - x') (x - x')^\top \right),\end{aligned}$$

with $M = A^\top A \succ 0$.

- Direct edge inference is possible if, for example,

$$d_\phi(x_i, x_j) \begin{cases} \leq \gamma - 1 & \text{for } x_i \sim x_j, \\ \geq \gamma + 1 & \text{for } x_i \not\sim x_j. \end{cases}$$

Problem formulation

- In the spirit of SVM, this suggests the following optimization problem:

$$\begin{aligned} \text{Minimize} \quad & \|M\|_{Fro}^2 + C \sum_{(i,j)} \zeta_{i,j} \\ \text{subject to} \quad & \zeta_{i,j} \geq 0, \quad \forall (i,j) \in \mathcal{T} \\ & d_M(x_i, x_j) \leq \gamma - 1 + \zeta_{i,j}, \quad i \sim j \\ & d_M(x_i, x_j) \geq \gamma + 1 - \zeta_{i,j}, \quad i \not\sim j \\ & M \succeq 0. \end{aligned}$$

- Similar to Tsang and Kwok (2003)

Representation of the solution

Lemma

The solution \hat{M} of the previous problem can be expanded as:

$$\hat{M} = \sum_{(i,j) \in \mathcal{T}} \alpha_{i,j} (\mathbf{x}_i - \mathbf{x}_j) (\mathbf{x}_i - \mathbf{x}_j)^\top,$$

where $\alpha_{i,j}$ are real number associated to the training pairs $(i, j) \in \mathcal{T}$.

Remarks

- The “classical” representer theorem can not be applied here because of the $M \succeq 0$ constraint. A slight extension is required.
- Plugging back into the problem leads to a **SDP** problem
- The dimension of the SDP is $2|\mathcal{T}| + 1 \implies$ **computational issues**.

Representation of the solution

Lemma

The solution \hat{M} of the previous problem can be expanded as:

$$\hat{M} = \sum_{(i,j) \in \mathcal{T}} \alpha_{i,j} (x_i - x_j) (x_i - x_j)^\top,$$

where $\alpha_{i,j}$ are real number associated to the training pairs $(i, j) \in \mathcal{T}$.

Remarks

- The “classical” representer theorem can not be applied here because of the $M \succeq 0$ constraint. A slight extension is required.
- Plugging back into the problem leads to a **SDP** problem
- The dimension of the SDP is $2|\mathcal{T}| + 1 \implies$ **computational issues**.

Relaxation

- Relax the constraint $M \succeq 0$

- The problem becomes:

$$\begin{aligned} \text{Minimize} \quad & \|M\|_{Fro}^2 + C \sum_{(i,j)} \zeta_{i,j} \\ \text{subject to} \quad & \zeta_{i,j} \geq 0, \quad \forall (i,j) \\ & \langle M, D_{i,j} \rangle_{Fro} - \gamma \leq -1 + \zeta_{i,j}, \quad i \sim j \\ & \langle M, D_{i,j} \rangle_{Fro} - \gamma \geq 1 - \zeta_{i,j}, \quad i \not\sim j. \end{aligned}$$

with the notation

$$D_{i,j} = (x_i - x_j)(x_i - x_j)^T.$$

- This is a SVM over $d \times d$ matrices with training set $(D_{i,j})_{(i,j) \in \mathcal{T}}$

Relaxation

- Relax the constraint $M \succeq 0$
- The problem becomes:

$$\begin{aligned} \text{Minimize} \quad & \|M\|_{Fro}^2 + C \sum_{(i,j)} \zeta_{i,j} \\ \text{subject to} \quad & \zeta_{i,j} \geq 0, \quad \forall (i,j) \\ & \langle M, D_{i,j} \rangle_{Fro} - \gamma \leq -1 + \zeta_{i,j}, \quad i \sim j \\ & \langle M, D_{i,j} \rangle_{Fro} - \gamma \geq 1 - \zeta_{i,j}, \quad i \not\sim j. \end{aligned}$$

with the notation

$$D_{i,j} = (x_i - x_j)(x_i - x_j)^\top.$$

- This is a SVM over $d \times d$ matrices with training set $(D_{i,j})_{(i,j) \in \mathcal{T}}$

- Relax the constraint $M \succeq 0$
- The problem becomes:

$$\begin{aligned} \text{Minimize} \quad & \|M\|_{Fro}^2 + C \sum_{(i,j)} \zeta_{i,j} \\ \text{subject to} \quad & \zeta_{i,j} \geq 0, \quad \forall (i,j) \\ & \langle M, D_{i,j} \rangle_{Fro} - \gamma \leq -1 + \zeta_{i,j}, \quad i \sim j \\ & \langle M, D_{i,j} \rangle_{Fro} - \gamma \geq 1 - \zeta_{i,j}, \quad i \not\sim j. \end{aligned}$$

with the notation

$$D_{i,j} = (x_i - x_j)(x_i - x_j)^\top.$$

- This is a **SVM over $d \times d$ matrices** with training set $(D_{i,j})_{(i,j) \in \mathcal{T}}$

Metric learning pairwise kernel

Pairwise inner product

$$\begin{aligned}\langle D_{x_1, x_2}, D_{x_3, x_4} \rangle_{Fro} &= \text{Trace} \left((x_1 - x_2)(x_1 - x_2)^\top (x_3 - x_4)(x_3 - x_4)^\top \right) \\ &= \left((x_1 - x_2)^\top (x_3 - x_4) \right)^2 \\ &= \left(x_1^\top x_3 - x_1^\top x_4 - x_2^\top x_3 + x_2^\top x_4 \right)^2.\end{aligned}$$

Pairwise kernel

$$\begin{aligned}K_{MLPK}((x_1, x_2), (x_3, x_4)) \\ &= \left(K(x_1, x_3) - K(x_1, x_4) - K(x_2, x_3) + K(x_2, x_4) \right)^2.\end{aligned}$$

Metabolic network

- 769 vertices: **enzymes**
- 3702 edges : **catalyze successive reactions**
- 3702 negative pairs (randomly sampled)
- 5-fold CV, 3 repeats, parameter optimization over the training set

Data	MLPK		TPPK	
	Accuracy	AUC	Accuracy	AUC
Expression	77.8 ± 0.2	84.5 ± 0.1	76.7 ± 0.3	83.3 ± 0.2
Localization	63.9 ± 0.4	68.2 ± 0.4	62.3 ± 0.1	65.8 ± 0.4
Phylogenetic profile	79.8 ± 0.1	84.9 ± 0.2	78.4 ± 0.1	83.4 ± 0.4
Yeast two-hybrid	76.6 ± 0.2	82.0 ± 0.1	59.2 ± 0.1	65.1 ± 0.6
Sum	83.9 ± 0.4	90.9 ± 0.3	84.2 ± 0.5	91.1 ± 0.3

Co-complex network

- 797 vertices: **proteins**
- 3280 edges : **member of the same complex**
- 3280 negative pairs (randomly sampled)
- 5-fold CV, 3 repeats, parameter optimization over the training set

Data	MLPK		TPPK	
	Accuracy	AUC	Accuracy	AUC
Localization	76.5 ± 0.1	76.8 ± 0.1	79.6 ± 0.1	83.1 ± 0.1
Chip-chip	82.4 ± 0.3	89.7 ± 0.1	63.8 ± 0.1	67.9 ± 0.3
Pfam	92.2 ± 0.2	98.2 ± 0.1	85.5 ± 0.1	91.7 ± 0.2
PSI-BLAST	90.0 ± 0.3	97.3 ± 0.1	88.3 ± 0.1	93.6 ± 0.2

Conclusion and future work

Summary

- A kernel method for distance metric learning, with an objective function optimized for graph inference
- A relaxation that leads to a SVM with a particular kernel for pairs
- Encouraging experimental results

Future work

- Assess the effect of relaxation
- Integration of multiple pairwise kernels

Reference

J.-P. Vert, J. Qiu and W. S. Noble, Metric learning pairwise kernel for graph inference, *preprint arXiv q-bio.QM/0610040*, 2006.