Graph-driven features extraction from microarray data using diffusion kernels and kernel CCA

Jean-Philippe Vert
Ecole des Mines de Paris
Jean-Philippe.Vert@mines.org

Minoru Kanehisa
Bioinformatics Center, Kyoto University
kanehisa@kuicr.kyoto-u.ac.jp

Abstract

We present an algorithm to extract features from high-dimensional gene expression profiles, based on the knowledge of a graph which links together genes known to participate to successive reactions in metabolic pathways. Motivated by the intuition that biologically relevant features are likely to exhibit smoothness with respect to the graph topology, the algorithm involves encoding the graph and the set of expression profiles into kernel functions, and performing a generalized form of canonical correlation analysis in the corresponding reproducible kernel Hilbert spaces.

Function prediction experiments for the genes of the yeast *S. Cerevisiae* validate this approach by showing a consistent increase in performance when a state-of-the-art classifier uses the vector of features instead of the original expression profile to predict the functional class of a gene.

1 Introduction

Microarray technology (DNA chips) is quickly becoming a major data provider in the post-genomics era, enabling the monitoring of the quantity of messenger RNA present in a cell for several thousands genes simultaneously. By submitting cells to various experimental conditions and comparing the expression profiles of different genes, a better understanding of the regulation mechanisms and functions of each gene is expected. As a matter of fact, early experiments confirmed that many genes with similar function yield similar expression patterns [4], and systematic use of state-of-the-art machine learning classification algorithms highlighted the possibility of gene function prediction from microarray data, at least for some functional categories [2].

Independently of microarray technology, decades of research in molecular biology have characterized the roles played by many genes as catalyzing chemical reactions in the cell. This information has now been integrated into databases such as KEGG [8], where series of successive chemical reactions arranged into pathways are represented, together with the genes catalyzing them. In particular one can extract from such a database a graph of genes, where two genes are linked whenever they catalyze two successive reactions.

The question motivating this report is whether the knowledge of this graph can help improve the performance of gene function prediction algorithms based on microarray data

only. To this end we propose a graph-driven feature extraction process, based on the idea that expression patterns which correspond to actual biological events, such as the activation or inhibition of a particular pathway, are more likely to be shared by genes close to each other in the graph than non-relevant patterns. Our approach consists in translating this intuition as a regularized version of canonical component analysis between the genes mapped to two reproducible kernel Hilbert spaces, defined respectively by a diffusion kernel [9] on the graph and a linear kernel on the expression profiles. This formulation leads to a well-posed problem equivalent to a generalized eigenvector problem [1].

2 Problem formulation

The set of genes is represented by a discrete set $\mathcal{X}$ of cardinality $|\mathcal{X}| = n$. The set of expression profiles is a mapping $e : \mathcal{X} \to \mathbb{R}^p$, where p is the number of measurements and $e(x)$ is the expression profile of gene x. In the sequel we assume that the set of profiles has been centered, i.e., $\sum_{x \in \mathcal{X}} e(x) = 0$.

The graph of genes extracted from the pathway database is represented by a simple graph $\Gamma = (\mathcal{X}, \mathcal{E})$, with the genes as vertices. Our goal is use this graph to extract features from the expression profiles. To this end we formally define a feature to be a real-valued mapping on the set of genes $f : \mathcal{X} \to \mathbb{R}$, and we denote by $\mathcal{F} = \mathbb{R}^{\mathcal{X}}$ the set of possible features. The set of centered features is denoted by $\mathcal{F}_0 = \left\{ f \in \mathcal{F} : \sum_{x \in \mathcal{X}} f(x) = 0 \right\}$.

In particular linear features extracted from expression profiles $f_{e,v}$ are defined, for any $v \in \mathbb{R}^p$, by $f_{e,v}(x) = v' e(x)$, for any $x \in \mathcal{X}$ (here and often in the sequel we use matrix notations, where v is a column vector and v' its transpose). We call $\mathcal{G} \subset \mathcal{F}_0$ the set of linear features. The normalized variance of a linear feature is defined by:

$$\forall f_{e,v} \in \mathcal{G}, \quad V(f_{e,v}) = \frac{\sum_{x \in \mathcal{X}} f_{e,v}(x)^2}{||v||^2}. \tag{1}$$

It is a first indicator of the possible relevance of a linear vector. Indeed biological events such as the synthesis of new molecules usually require the coordinated actions of many proteins: they are therefore likely to have characteristic patterns in terms of gene expression which capture variation between the genes involved and the others, and should therefore have large variance. Linear features with a large normalized variance (1) are called *relevant* in the sequel, as opposed to *irrelevant* features. Relevant features can be extracted by PCA.

While the normalized variance (1) is an intrinsic property of the set of profiles, the knowledge of the graph Γ suggests another criterion to judge "good" features. As genes linked together in the graph are supposed to participate in successive reactions in the cell, it is likely that the activation/inhibition of a biochemical pathway has a characteristic expression pattern shared by clusters of genes in the graph. More globally, the graph defines a structure on the set of genes, and therefore a notion of *smoothness* for any feature $f \in \mathcal{F}$. A feature is called *smooth* if it varies slowly between adjacent nodes in the graph, and *rugged* otherwise. As just stated, features of interest are more likely to be smooth than other features.

We therefore end up with two criteria for extracting "good" features: they should simultaneously be *relevant* and *smooth*, the latter being defined with respect to the gene graph. One way to extract such features is to look for pairs of features, $(f_1, f_2) \in \mathcal{F} \times \mathcal{G}$, such that f_1 be smooth, f_2 be a relevant linear feature, and the correlation between f_1 and f_2 be as large as possible. The decoupling of the two criteria enables us to state the problem mathematically as follows.

Suppose we can define a smoothness functional $h_1 : \mathcal{F} \to \mathbb{R}^+$ for any feature, and a relevance functional $h_2 : \mathcal{G} \to \mathbb{R}^+$ for linear features, in such a way that lower values of

the functional h_1 (resp. h_2) correspond to smoother (resp. more relevant) features. Then the following optimization problem:

$$\max_{(f_1,f_2)\in\mathcal{F}_0\times\mathcal{G}} \frac{f_1' f_2}{\sqrt{f_1' f_1 + \delta h_1(f_1)}\sqrt{f_2' f_2 + \delta h_2(f_2)}}, \tag{2}$$

where $\delta > 0$ is a regularization parameter, is a way to extract smooth and relevant features. Irrelevance and ruggedness penalize any candidate pair through the functionals h_1 and h_2, and δ controls the trade-off between relevance and smoothness on the one hand, and correlation on the other hand. $\delta = 0$ amounts to finding f_1 and f_2 as correlated as possible (which is obtained by taking $f_1 = f_2$), while $\delta > 0$ forces f_1 to be relevant and f_2 to be smooth.

In order to turn (2) into an algorithm we remark that if h_1 and h_2 can be expressed as norms in reproducible kernel Hilbert spaces (RKHS, see Section 3), then (2) takes the form of a generalization of canonical correlation analysis (CCA) known as kernel-CCA [1], which is equivalent to a generalized eigenvector problem. Let us therefore show how to build two RKHS on the set of genes whose norms are smoothness (Section 4) and relevance (Section 5) functionals, respectively.

3 Reproducible kernel Hilbert spaces and smoothness functionals

Let us briefly review basic properties of RKHS relevant for the sequel. The reader is referred to [12, 14] for more details.

Let $K : \mathcal{X}^2 \to \mathbb{R}$ be a Mercer kernel in the sense that the matrix $K = (K_{x,y})_{(x,y)\in\mathcal{X}^2}$ be symmetric positive semidefinite. Let $\mathcal{H} \subset \mathcal{F}$ be the linear span of $\{K(x,.), x \in \mathcal{X}\}$, and consider a decomposition of K as:

$$K = \sum_{i=1}^{n} \lambda_i \phi_i \phi_i', \tag{3}$$

where $0 \leq \lambda_1 \leq \ldots \leq \lambda_n$ are the eigenvalues of K and the set $(\phi_1, \ldots, \phi_n) \in \mathcal{F}^n$ is an associated orthonormal basis of eigenvectors in $L^2(\mathcal{X})$. The decomposition of any $f \in \mathcal{H}$ on this basis can be expressed as $f = \sum_{i=r+1}^{n} a_i \phi_i$, where r is the multiplicity of 0 as an eigenvalue. An inner product can be defined in $\mathcal{H}$ as follows:

$$\left\langle \sum_{i=r+1}^{n} a_i \phi_i, \sum_{i=r+1}^{n} b_i \phi_i \right\rangle_{\mathcal{H}} = \sum_{i=r+1}^{n} \frac{a_i b_i}{\lambda_i}. \tag{4}$$

The resulting Hilbert space $\mathcal{H}$ is called a reproducing kernel Hilbert space, due to the following reproducing property:

$$\forall (x,x') \in \mathcal{X}^2, \quad \langle K(.,x), K(.,x') \rangle_{\mathcal{H}} = K(x,x'). \tag{5}$$

The inner product in $\mathcal{H}$ can be easily expressed in a dual form as follows. Each $f \in \mathcal{H}$ can be decomposed as $f(.) = \sum_{x\in\mathcal{X}} \alpha(x) K(x,.)$, where α is unique up to the addition of an element of the null space of K and is called the dual coordinate of f. In a matrix form, this reads $f = K\alpha$, and using (5) one can easily check that the inner product between two features $(f,g) \in \mathcal{H}^2$ with dual coordinates $(\alpha,\beta) \in \mathcal{F}^2$ respectively is given by:

$$\langle f, g \rangle_{\mathcal{H}} = \sum_{(x,y)\in\mathcal{X}^2} \alpha(x)\beta(y)K(x,y) = \alpha' K \beta. \tag{6}$$

In particular the $\mathcal{H}$-norm of a feature $f \in \mathcal{H}$ with dual coordinates $\alpha \in \mathcal{F}$ is given by:

$$||f||_{\mathcal{H}}^2 = \alpha' K \alpha, \tag{7}$$

and the inner product between two features $(f, g) \in \mathcal{H}^2$ with dual coordinates $(\alpha, \beta) \in \mathcal{F}^2$ in the original space $L^2(\mathcal{X})$ can also be expressed in dual form:

$$f'g = \sum_{x \in \mathcal{X}} f(x)g(x) = \alpha' K^2 \beta. \tag{8}$$

When $\mathcal{X}$ is a subspace of $\mathbb{R}^N$ then it is known that the norm in the RKHS defined by several popular kernels such as the Gaussian radial basis kernel are smoothing functionals, in the sense that larger values of $||f||_{\mathcal{H}}$ correspond to functions f with more energy at high frequency in their Fourier decomposition. This fact has been much exploited e.g. in regularization theory [14, 5], and we now adapt it to the discrete setting.

4 Smoothness functional on a graph

A natural way to quantify the smoothness of a feature on a graph is by its energy at high frequency, as computed from its Fourier transform. Fourier transforms on graphs is a classical tool of spectral graph analysis [3, 11] which we briefly recall now. Let A be the $n \times n$ adjacency matrix of the graph Γ ($A_{x,y} = 1$ if there is an edge between x and y, 0 otherwise) and D the diagonal matrix of vertex degrees. Then the $n \times n$ matrix $L = D - A$ is called the *Laplacian* of Γ, and is known to share many properties with the continuous Laplacian [11]. It is symmetric, semidefinite positive, and singular. The eigenvector $(1, \ldots, 1)$ belongs to the eigenvalue $\lambda_1 = 0$, whose multiplicity is equal to the number of connected components of Γ.

Let us denote by $0 = \lambda_1 \leq \ldots \leq \lambda_n$ the eigenvalues of L and $\{\phi_i, i = 1, \ldots, n\}$ an orthonormal set of associated eigenvectors. This basis is a discrete Fourier basis [3], and it is known that ϕ_i oscillates more and more on the graph as i increases. The Fourier decomposition of any feature $f \in \mathcal{F}$ is the expansion in terms of this basis:

$$f = \sum_{i=1}^{n} \hat{f}_i \phi_i, \tag{9}$$

where $\hat{f}_i = \phi_i' f$ and $\hat{f} = \left(\hat{f}_1, \ldots, \hat{f}_n\right)$ is called the *discrete Fourier transform* of f.

For any monotonic decreasing mapping $\zeta : \mathbb{R}^+ \to \mathbb{R}^+ \backslash \{0\}$, let us now consider the function $K_\zeta : \mathcal{X}^2 \to \mathbb{R}$ defined by:

$$\forall (x, y) \in \mathcal{X}^2, \qquad K_\zeta(x, y) = \sum_{i=1}^{n} \zeta(\lambda_i)\phi_i(x)\phi_i(y). \tag{10}$$

The mapping ζ being assumed to take only positive values, the matrix K_ζ is definite positive and is therefore a Mercer kernel on the set $\mathcal{X}$. The corresponding RKHS is the set of features $\mathcal{F}$, with norm given by:

$$\forall f \in \mathcal{F}, \quad ||f||^2_{K_\zeta} = \sum_{i=1}^{n} \frac{\hat{f}_i^2}{\zeta(\lambda_i)}. \tag{11}$$

As i increases, λ_i increases so $\zeta(\lambda_i)$ decreases. As a result the norm (11) has a higher value on features which have a lot of energy at high frequency, and is therefore a natural smoothing functional.

An example of valid ζ function with rapid decay is the exponential $\zeta(x) = e^{-\tau x}$, where τ is a parameter. In that case we recover the *diffusion kernel* introduced and discussed in [9]. Considering other mapping ζ would be beyond the scope of this report, so we restrict ourselves to this diffusion kernel in the sequel. Observe that it can be expressed using the matrix exponential as $K_\zeta = \exp(-\tau L)$.

5 Relevance functional

If $v \in \mathbb{R}^p$ has a projection v_0 onto the linear span of $\{e(x), x \in \mathcal{X}\}$ then $f_{e,v} = f_{e,v_0}$. As a result the set of linear features $\mathcal{G}$ can be parametrized by directions of the form $v = \sum_{x \in \mathcal{X}} \beta(x) e(x)$, where $\beta \in \mathcal{F}$ is called the dual coordinate of v and is defined up to the addition of an element of the null space of the Gram matrix $K_{x,y} = e(x)'e(y)$. The RKHS $\mathcal{H} \subset \mathcal{F}$ associated with this semidefinite positive matrix consists of the set of features of the form $f(.) = \sum_{x \in \mathcal{X}} \beta(x) K(x,.) = f_{v,e}$, where $v = \sum_{x \in \mathcal{X}} \beta(x) e(x)$. In other words this is exactly the set of linear features, $\mathcal{H} = \mathcal{G}$.

The variance of a feature $f \in \mathcal{G}$ can be expressed by (1), (6) and (8) as follows:

$$V(f_{e,v}) = \frac{\sum_{x \in \mathcal{X}} f_{e,v}(x)^2}{||v||^2} = \frac{\beta' K^2 \beta}{\beta' K \beta} = \frac{||f_{e,v}||_{L^2(\mathcal{X})}}{||f_{e,v}||_{\mathcal{H}}}.$$

As a result, a natural relevance functional to balance the term $||f||_{L^2(\mathcal{X})}$ in (2) is the norm in the RKHS: $h_2(f_{e,v}) = ||f_{e,v}||_{\mathcal{H}}$, where $\mathcal{H}$ is the RKHS associated with the linear kernel $K(x,y) = e(x)'e(y)$.

6 Extracting smooth correlations

Let $K_1 = \exp(-\tau L)$ denote the diffusion kernel and K_2 denote the linear kernel $K_2(x,y) = e(x)'e(y)$, with associated RKHS $\mathcal{H}_1$ and $\mathcal{H}_2$ respectively. Taking $h_1(f) = ||f||_{\mathcal{H}_1}$ as a smoothness function for any $f \in \mathcal{F}$, and $h_2(f) = ||f||_{\mathcal{H}_2}$ as a relevance functional for any linear feature $f \in \mathcal{G}$, we can express the maximization Problem (2) in a dual form as:

$$\max_{(\alpha,\beta) \in \mathcal{F}^2} \gamma(\alpha,\beta) \triangleq \frac{\alpha' K_1 K_2 \beta}{\left(\alpha' \left(K_1^2 + \delta K_1\right) \alpha\right)^{\frac{1}{2}} \left(\beta' \left(K_2^2 + \delta K_2\right) \beta\right)^{\frac{1}{2}}}. \tag{12}$$

At first sight it seems that (12) is the dual formulation of an optimization over $(f_1, f_2) \in \mathcal{H}_1 \times \mathcal{H}_2 = \mathcal{F} \times \mathcal{G}$, and not $\mathcal{F}_0 \times \mathcal{G}$ as in (2). However it can be checked that any solution of (12) is in fact in $\mathcal{F}_0 \times \mathcal{G}$. Indeed the numerator remains unchanged when a constant function is added to $f_1 = K_1 \alpha \in \mathcal{F}$, while both $||f_1||_{L^2(\mathcal{X})}$ and $||f_1||_{\mathcal{H}_1}$ are minimized when f has mean 0 (for the latter case, this results from the fact that the constant vector is an eigenvector of the diffusion kernel, so the norm defined by (4) is minimized when the corresponding projection of f, namely its average, is null).

Formulated as (12) the problem appears to be a generalization of canonical correlation analysis (CCA) known as kernel-CCA, discussed in [1]. In particular Bach and Jordan show that (α, β) is a solution of (12) if and only if it satisfies the following generalized eigenvalue problem:

$$\begin{pmatrix} 0 & K_1 K_2 \\ K_2 K_1 & 0 \end{pmatrix} \begin{pmatrix} \alpha \\ \beta \end{pmatrix} = \rho \begin{pmatrix} K_1^2 + \delta K_1 & 0 \\ 0 & K_2^2 + \delta K_2 \end{pmatrix} \begin{pmatrix} \alpha \\ \beta \end{pmatrix} \tag{13}$$

with ρ the largest possible. Moreover, solving (13) provides a series of pairs of features $\{(\alpha_i, \beta_i), i = 1, \ldots, \bar{n}\}$, where $\bar{n} = \min(n,p)$, with decreasing values of $\gamma(\alpha_i, \beta_i)$ for which the gradient $\nabla_{\alpha,\beta} \gamma$ is null, equivalent to the extraction of successive canonical directions with decreasing correlation in classical CCA. The resulting features $f_{1,i} = K_1 \alpha_i$ and $f_{2,i} = K_2 \beta_i$ are therefore a set of features likely to have decreasing biological relevance when i increases, and are the features we propose to extract in this report.

As discussed in [1] we regularize the problem (13) by adding $\delta^2/4$ on the diagonal of the matrix on the right-side, to be able to perform the Cholesky decomposition necessary to

solve this problem. Hence we end up with the following problem:

$$\begin{pmatrix} 0 & K_1K_2 \\ K_2K_1 & 0 \end{pmatrix} \begin{pmatrix} \alpha \\ \beta \end{pmatrix} = \rho \begin{pmatrix} (K_1 + \delta' I)^2 & 0 \\ 0 & (K_2 + \delta' I)^2 \end{pmatrix} \begin{pmatrix} \alpha \\ \beta \end{pmatrix}, \quad (14)$$

where $\delta' = \delta/2$. If (α, β) is an generalized eigenvector solution of (14) belonging to the generalized eigenvalue ρ, then $(-\alpha, \beta)$ belong to $-\rho$. As a result the spectrum of (14) is symmetric : $(\rho_1, -\rho_1, \ldots, \rho_n, -\rho_n)$ with $\rho_1 \geq \ldots \geq \rho_n$, $\rho_i = 0$ for $i > p$.

7 Experiments

We extracted from the LIGAND database of chemical compounds of reactions in biological pathways [6] a graph made of 774 genes of the budding yeast *S. Cerevisiae*, linked through 16,650 edges, where two genes are linked when they have the possibility to catalyze two successive reactions in the LIGAND database (i.e, two reactions such that the main product of the first one be the main substrate of the second one). Expression data were collected from the Stanford Microarray Database [13]. Concatenating several publicly available data, we ended up with 330 measurements for 6075 genes of the yeast, i.e., almost all its known or predicted genes. Following [4, 2] we work with the normalized logarithm of the ratio of expression levels of the genes between two experimental conditions. The functional classes of the yeast genes we consider are the one defined by the January 10, 2002 version of the Comprehensive Yeast Genome Database (CYGD) [10], which is a comprehensive classification of 3,936 genes into 259 categories.

The 669 genes in the gene graph with known expression profiles were first used to perform the feature extraction process described in this report. The resulting linear features were then extracted from the expression profiles of the disjoint set of 2,688 genes which are in the CYGD functional catalogue but not in the pathway database. We then performed functional classification experiments on this set of 2,688 genes, using either the profiles themselves or the features extracted. All functional classes with more than 20 members in this set were tested (which amount to 115 categories).

Experiments were carried out with *SVM Light* [7], a public and free implementation of SVM. All vectors were scaled to unit length before being sent to the SVM, and all SVM use a radial basis kernel with unit width, i.e., $k(x, y) = \exp(-||x - y||^2)$. The trade-off parameter between training error and margin error was set to its default value (1 in that case), and the cost of errors on positive and and negative examples were adjusted to have the same total.

Preliminary experiments to tune the two parameters of the algorithm, namely the width of the diffusion kernel τ and the regularization parameter δ, showed that $\tau = 1$ and $\delta = 0.001$ provide good performances. For these values we first tested whether there exists an optimal number of features to be extracted for optimal gene function prediction. Figure 1 shows the performance of SVM using different numbers of features, in terms of ROC index averaged over all 115 classes. The ROC index is the area under the curve of false negative vs true positive, normalized to 100 for a perfect classifier and 50 for a random classifier. For each category the ROC index was averaged over 10 random splitting of the data into training and test set, in the proportion $80/20$. It appears that the more features are included, the better the performance averaged over all categories. A more precise analysis of the different classes shows however that some classes don't follow the average trend and are better predicted by a smaller number of features, as shown on Figure 2 for 5 categories best predicted by less than 100 features. Finally Figure 3 compares, for each of the 115 categories, the ROC index for a SVM using the original expression profiles with a SVM using the vectors of 330 features. It demonstrates that the representation of genes as vectors of features helps improve the performance of SVM (the ROC index averaged over all categories increases

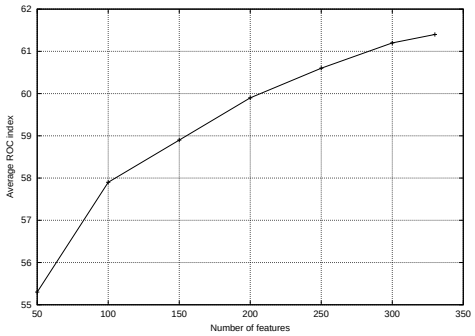

Figure 1: ROC index averaged over 115 categories, for various number of features

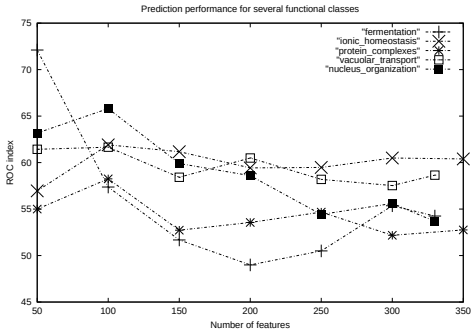

Figure 2: ROC index for 5 functional categories, for various number of features

from 54.6 to 61.4). The difference is especially important for classes such as heavy metal ion transporters (83.5 vs 55.2), ribosome biogenesis (94.6 vs 70.9), protein synthesis (84.3 vs 61.6) or morphogenesis (64, 7 vs 44.3)

8 Discussion and Conclusion

Results reported in the previous section are encouraging for at least two reasons. First of all, the performance reached for some classes such as heavy ion metal transporters shows that a ROC above 80% can be expected for several classes. Second, while many classes are apparently not learned by the SVM based on expression profiles (ROC around 50), the ROC based on extracted features of the same classes is around 60. This shows that there is hope to be able to predict more functional classes than previously thought [2] from microarray data, which is a good news since the amount of microarray data is expected to explode in the coming years.

The method presented in this paper can be seen as an attempt to explore the possibilities of data mining and analysis provided by kernel methods. Few studies have used kernel methods other than SVM, and have used kernels other than Gaussian or polynomial kernels. In this report we tried to show how "exotic" kernels such as the diffusion kernel, and "exotic" methods such as kernel-CCA, can be adapted to particular problems, graph-driven feature extraction in our case. Exploring other possibilities of kernel methods in the data-rich field of computational biology is among our future plans.

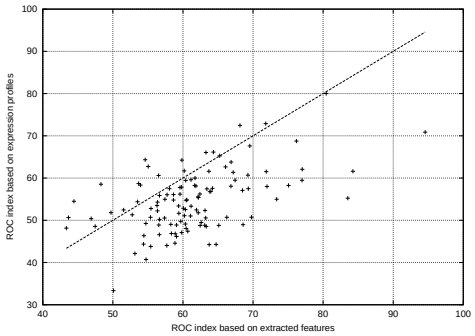

Figure 3: ROC index of a SVM classifier based on expression profiles (y axis) or extracted features (x axis). Each point represents one functional category.

References

[1] F. R. Bach and M. I. Jordan. Kernel independent component analysis. *Journal of Machine Learning Research*, 3:1–48, 2002.

[2] Michael P. S. Brown, William Noble Grundy, David Lin, Nello Cristianini, Charles Walsh Sugnet, Terence S. Furey, Jr. Manuel Ares, and David Haussler. Knowledge-based analysis of microarray gene expression data by using support vector machines. *Proc. Natl. Acad. Sci. USA*, 97:262–267, 2000.

[3] Fan R.K. Chung. *Spectral graph theory*, volume 92 of *CBMS Regional Conference Series*. American Mathematical Society, Providence, 1997.

[4] Michael B. Eisen, Paul T. Spellman, Patrick O. Brown, and David Botstein. Cluster analysis and display of genome-wide expression patterns. *Proc. Natl. Acad. Sci. USA*, 95:14863–14868, Dec 1998.

[5] Frederico Girosi, Michael Jones, and Tomaso Poggio. Regularization theory and neural networks architectures. *Neural Computation*, 7(2):219–269, 1995.

[6] S. Goto, Y. Okuno, M. Hattori, T. Nishioka, and M. Kanehisa. LIGAND: database of chemical compounds and reactions in biological pathways. *Nucleic Acid Research*, 30:402–404, 2002.

[7] Thorsten Joachims. Making large-scale svm learning practical. In B. Schölkopf, C. Burges, and A. Smola, editors, *Advances in Kernel Methods - Support Vector Learning*, pages 169–184. MIT Press, 1999.

[8] M. Kanehisa, S. Goto, S. Kawashima, and A. Nakaya. The KEGG databases at GenomeNet. *Nucleic Acid Research*, 30:42–46, 2002.

[9] R. I. Kondor and J. Lafferty. Diffusion kernels on graphs and other discrete input. In *ICML 2002*, 2002.

[10] H.W. Mewes, D. Frishman, U. Güldener, G. Mannhaupt, K. Mayer, M. Mokrejs, B. Morgenstern, M. Münsterkoetter, S. Rudd, and B. Weil. MIPS: a database for genomes and protein sequences. *Nucleic Acid Research*, 30(1):31–34, 2002.

[11] B. Mohar. Some applications of laplace eigenvalues of graphs. In G. Hahn and G. Sabidussi, editors, *Graph Symmetry: Algebraic Methods and Applications*, volume 497 of *NATO ASI Series C*, pages 227–275. Kluwer, Dordrecht, 1997.

[12] S. Saitoh. *Theory of reproducing Kernels and its applications*. Longman Scientifi c & Technical, Harlow, UK, 1988.

[13] G. Sherlock, T. Hernandez-Boussard, A. Kasarskis, G. Binkley, J.C. Matese, S.S. Dwight, M. Kaloper, S. Weng, H. Jin, C.A. Ball, M.B. Eisen, and P.T. Spellman. The stanford microarray database. *Nucleic Acid Research*, 29(1):152–155, Jan 2001.

[14] G. Wahba. *Spline Models for Observational Data*, volume 59 of *CBMS-NSF Regional Conference Series in Applied Mathematics*. SIAM, Philadelphia, 1990.