TOPOLOGICAL EXPANSION FOR SMOOTHING THE KERNEL BASED ON BIOMEDICAL AND MICROARRAY DATA

Vilen Jumutc
Riga Technical University
Meza 1/4, LV-1658 Riga, Latvia

Pawel Zayakin
Latvian BioMedical Research & Study Center
Ratsupites 1, LV-1067 Riga, Latvia

ABSTRACT

In this paper we propose a new approach in the development of kernel functions (kernels) for Support Vector Machines (SVMs) that could be applied in practical biomedicine as robust and sufficiently smooth classification and regression toolbox for cancer outcome prediction and diagnostic decision support commonly based on microarray, image and voice data. Recent major interest in biomedicine is connected with microarray data and its possible diagnostic and prognostic value. Nevertheless performed SVM classification frequently is based solely on kernels that represent very simple and well-known similarity concepts, i.e. inner products and their expansions to some higher dimensional Hilbert space. On the other hand topological expansion of frequently used RBF (Gaussian) and linear kernels is the main intent of this paper. Each gene or antigen "printed" on microarray chip could represent not only expression level but up- or down-regularity present in examined sample with respect to other involved (anti)genes. This information carefully rewritten could represent each sample topology and could serve as the topological expansion of every kernel. The experimental evaluations are performed on different biomedical datasets and verify that proposed kernel improves performance on purely conditioned and even very small training sets.

KEYWORDS

SVM, kernel methods, MKL, topological expansion.

1. INTRODUCTION

In support vector machines (SVMs) optimal kernel selection is a crucial condition for classification success and good generalization on unseen data. Merely generalization success is connected with the right similarity measure encoded in the kernel function and thus depends on some properties of feature space or appropriate expansion of inner product represented by selected kernel in enriched Hilbert space. From this point of view it is clear that for standard linear, RBF etc. kernels proper cross-sample normalization and input space scaling is needed for successful SVM learning (large-margin optimization) and smaller parameters' subspace that should be searched by cross-validation or some more sophisticated Multiple Kernel Learning (MKL) approach in order to find optimal "tuning" parameters (for instance bandwidth parameter of RBF kernel).

Mentioned above constraints on SVM classification require very careful and even per-sample approach in preliminary data normalization for biomedical and microarray data sources that usually happens because of high variability of a signal strength on microarray chips or poorly conditioned experimental environment.

Our proposed kernel doesn't require global cross-sample normalization and works solely with ranking information available for each attribute in the sample. This property basically helps to avoid expensive and time-consuming normalization and provides classification with robust and even more accurate estimation of similarity without even proper rescaling of input space.

In general we train and test standard RBF and our proposed kernel within simpleMKL framework (Rakotomamonjy et al., 2008) in order to avoid time-expensive cross-validation and provide more accurate estimation of "tuning" parameters.

2. BACKGROUND

In this section we briefly describe our primary data source, SVM and MKL basics.

2.1 Microarray Data Source

Circulating autoantibodies against tumour-derived proteins have been observed in the most if not all cancer patients therefore they seem to be very attractive targets for the development of noninvasive serological tests for the diagnosis or early detection of cancer. Moreover, the induction of tumour-specific B cell responses by immunotherapy and standard treatments such as radiation and hormone therapies has been observed suggesting that autoantibodies potentially could be exploited as biomarkers of response to therapy. With a phage-display library derived from melanoma, gastric and prostate cancer tissues, we developed and used phage protein microarrays from a set of 1229 different serum-reactive phage clones to analyze serum samples from 172 patients with gastric cancer, 167 patients with melanoma cancer, 52 patients with prostate cancer and 147 samples from healthy control group (healthy donors, HD) as it was developed and described in (Kalnina, 2008a; Kalnina, 2008b). Further each subset of cancer specific patients was examined versus all HD samples in classification trials of SVM with proposed and RBF kernels.

2.2 Support Vector Machines

Support Vector Machines are based on the concept of decision planes that define decision boundaries using statistical learning theory (Vapnik, 1995). Support Vector Machine (SVM) models are a close cousin to classical multilayer ANN. Using a kernel function, SVM is an alternative training method for polynomial, radial basis function and multi-layer perceptron classifiers in which the weights of the network are found by solving the quadratic programming problem with linear constraints, rather than by solving a non-convex, unconstrained minimization problem as in standard neural network training.

A basic formalization of hard-margin SVM can be stated as follows: minimize $\|w\|^2$ (2-norm of margin γ between two hyperplanes)

$$s.t. \quad \gamma = \frac{2}{\|w\|}$$

with respect to the following constraints:

$$\mathrm{y}_\mathrm{i}\left(w' x_i + b\right) \geq 1 \quad i = \overline{1,n},\ x_i \in R^d,\ y_i \in \{-1,1\}$$

or in the case of soft-margin SVM:

$$\mathrm{y}_\mathrm{i}\left(w' x_i + b\right) \geq 1 - \xi_\mathrm{i} \quad i = \overline{1,n},\ x_i \in R^d,\ y_i \in \{-1,1\}, \xi_i \geq 0$$

Finally the quadratic programming problem and the target function (assuming hard-margin case) with respect to all constraints can be generalized by the following Lagrangian and KKT conditions:

$$\min_w \left(\max_{\lambda>0} L(w,\lambda)\right) \quad L(w,\lambda) = \frac{1}{2}\|w\|^2 - \sum_{i=1}^{n} \lambda_i \left(\mathrm{y}_\mathrm{i}\left(w' x_i + b\right) - \mathbf{1}\right)$$

Karush–Kuhn–Tucker (KKT) conditions:

$$\frac{\partial L(w,b,\lambda)}{\partial w} = \frac{\partial L(w,b,\lambda)}{\partial b} = 0$$

$$\lambda_i \left(\mathrm{y}_\mathrm{i}\left(w' x_i + b\right) - \mathbf{1}\right) = 0 \quad i = \overline{1,n}$$

where the last one is complementary slackness condition.

After taking into consideration KKT conditions final SVM dual representation (Lagrangian) is expressed only in terms of λ_i. Thus it can be effectively optimized using any off-shell linear optimizer that supports constraint adaptation:

$$\max_{\lambda} \sum_{i=1}^{l} \lambda_i - \frac{1}{2} \sum_{i=1}^{l} \sum_{j=1}^{l} \lambda_i \lambda_j y_i y_j \langle \mathbf{x}_i, \mathbf{x}_j \rangle, \quad s.t. \quad \lambda_i \geq 0, \quad \sum_{i=1}^{l} y_i \lambda_i = 0$$

In the case of soft-margin SVM constraints on dual variables are slightly modified in order to include trade-off parameter C in its primal formulation:

$$\min_{w} \frac{1}{2} \|w\|^2 + C \sum_{i} \xi_i$$

In the mentioned soft-margin case the Lagrangian with respect to dual variables (after taking into consideration corresponding KKT conditions) is the same as in hard-margin case but constraints on dual variables include trade-off parameter:

$$C \geq \lambda_i \geq 0, \quad \sum_{i=1}^{l} y_i \lambda_i = 0$$

Here we can see that target function uses inner product between input vectors and with the help of "kernel trick" it is possible to express this product in terms of higher dimensional mapping of vectors $\phi(x)$ that allows determination of suitable separating hyperplane in this higher dimensional feature space even when in input space data is not linearly separable. Thus in terms of this mapping to feature space our target functions is expressed as follows:

$$\max_{\lambda} \sum_{i=1}^{l} \lambda_i - \frac{1}{2} \sum_{i=1}^{l} \sum_{j=1}^{l} \lambda_i \lambda_j y_i y_j \langle \phi(\mathbf{x}_i), \phi(\mathbf{x}_j) \rangle$$

And as already mentioned above with the help of "kernel trick" we could define kernel function $K(x, x')$ or simply kernel that expresses inner product between input vectors in higher dimensional feature space avoiding explicit transformation of input vectors to feature space. Thus in terms of predefined kernel our target function is expressed as follows:

$$\max_{\lambda} \sum_{i=1}^{l} \lambda_i - \frac{1}{2} \sum_{i=1}^{l} \sum_{j=1}^{l} \lambda_i \lambda_j y_i y_j K(x_i, x_j)$$

Finally classification of a new sample is derived by: $d = sign(w' x + b)$

2.3 Multiple Kernel Learning

Multiple Kernel Learning (MKL) aims at simultaneously learning a kernel and the associated predictor in general SVM context. For the SVM, an efficient and general MKL algorithm, based on semi-infinite linear programming (SILP) has been proposed recently and as it has been shown in related works could be easily applied to learning optimal kernel parameters without facing the cross-validation for large biomedical problems (Sonnenburg et al., 2006).

Recent applications of MKL have clearly proven that using multiple kernels instead of a single one can enhance the interpretability of the decision function and improve performances (Lanckriet et al., 2004). In such cases, a convenient approach is to consider that the kernel $K(x, x')$ is actually a convex combination of basis kernels:

$$K(x, x') = \sum_{m=1}^{M} d_m K_m(x, x'), \quad d_m \geq 0, \quad \sum_{m=1}^{M} d_m = 1$$

where M is the total number of kernels. Each basis kernel K_m may either use the full set of variables describing x or subsets of variables extracted from different data sources (Lanckriet et al., 2004). Alternatively, the kernels K_m can simply be standard kernels (such as Gaussian or RBF kernels) with different parameterization. Within this framework, the problem of data representation through the kernel is then transferred to the choice of optimal weights d_m that minimizes the MKL objective function (Bach et al., 2004).

In our work we have used MKL method to perform general classification task and estimate optimal bandwidth parameter (γ) of the basis RBF kernel being employed in our topological kernel as well.

3. TOPOLOGICAL KERNEL

In this section we propose new topological kernel as well as present some preliminary generalization bounds obtained for this kernel in terms of Rademacher complexity.

3.1 General Definition

Our proposed kernel uses rank information available for each attribute in the sample. This information is acquired by introducing so-called topological measure of every attribute that has its own relative disposition in the sample that doesn't depend on other samples and can be completely regarded as an ordinal ranking of a "signal" encoded by this attribute within each sample. This topological measure can be viewed as two separate quantifiers that account up- and down-regularity of selected attribute and corresponding "signal" in comparison to other attributes. The proposed up-regulation based topological measure for *i-th* attribute of each sample is given as follows:

$$\Omega_{up}\left(x^{(i)}\right)=\sum_{j=1}^{m} I\left(x^{(j)} > x^{(i)}\right)$$

Hereby I is an indicator function and m is dimensionality of classification problem. Similarly down-regulation based topological measure for *i-th* attribute of each sample is given as follows:

$$\Omega_{down}\left(x^{(i)}\right)=\sum_{j=1}^{m} I\left(x^{(j)} \le x^{(i)}\right)$$

Finally newly proposed kernel that incorporates mentioned topological measures of two independent samples can be viewed as a dot product of such quantifiers among all attributes. Linear representation of this kernel can be obtained as follows:

$$K_{linear}\left(x_i, x_j\right)=\delta \cdot \sum_{k=1}^{m} \Omega_{up}\left(x_i^{(k)}\right) \cdot \Omega_{up}\left(x_j^{(k)}\right)+\Omega_{down}\left(x_i^{(k)}\right) \cdot \Omega_{down}\left(x_j^{(k)}\right)$$

Hereby δ is some normalization factor that prevents the value of newly proposed kernel function to become too large.

Analogically RBF kernel in the terms of the newly proposed kernel function could be expressed as expansion of previously defined linear kernel to some highly dimensional Hilbert space:

$$K_{RBF}\left(x_i, x_j\right)=\exp\left(-\gamma\left\|x_i-x_j\right\|^2\right)=\exp\left(-\gamma\left(K_{linear}\left(x_i, x_i\right)-2 \cdot K_{linear}\left(x_i, x_j\right)+K_{linear}\left(x_j, x_j\right)\right)\right)$$

Hereby γ is bandwidth parameter of RBF kernel. Finally self-normalized kernel expansion for mentioned above topological dot product can be inferred following the assumption that every attribute's topological measure is surely upper bounded by the dimension of the classification problem. Thus we could easily represent self-normalized up-regulation and down-regulation based topological measure for *i-th* attribute of each sample as follows:

$$\Omega_{up}\left(x^{(i)}\right)=\frac{1}{m}\sum_{j=1}^{m} I\left(x^{(j)} > x^{(i)}\right) \qquad \Omega_{down}\left(x^{(i)}\right)=\frac{1}{m}\sum_{j=1}^{m} I\left(x^{(j)} \le x^{(i)}\right)$$

Hereby m is input dimensionality of classification problem. Another useful expansion of proposed kernel comes from the fact that sometimes sample topology is more rigidly correlated not with sample input dimensionality, but with the number of unique attribute values characterizing that sample. In the sense of this assumption we have inferred another useful expansion of topological measure for down- and up-regularity of *i-th* attribute:

$$\Omega_{up}\left(x^{(i)}\right)=\frac{1}{|\tau|}\sum_{j=1}^{m} I\left(\tau_j > x^{(i)}\right) \qquad \Omega_{down}\left(x^{(i)}\right)=\frac{1}{|\tau|}\sum_{j=1}^{m} I\left(\tau_j \le x^{(i)}\right)$$

Hereby τ is a vector of all possible unique "signals" within sample x. Consequently topological measures for each attribute of the sample x are self-normalized by the underlying topology of the whole sample x. In our experimental phase we commonly used the last version of topological measures conforming to the following topological kernel w/o normalization factor:

$$K_{linear}\left(x_i, x_j\right)=\sum_{k=1}^{m} \Omega_{up}\left(x_i^{(k)}\right) \cdot \Omega_{up}\left(x_j^{(k)}\right)+\Omega_{down}\left(x_i^{(k)}\right) \cdot \Omega_{down}\left(x_j^{(k)}\right)$$

3.2 Generalization Bounds for Topological Kernel

In this section we propose effective SVM generalization bound in terms of Rademacher complexity and specificity of proposed topological kernel. In (Bartlett and Mendelson, 2002) authors yielded effective upper bound on generalization error using Rademacher complexity

$$R_n(F) = \frac{2}{n} \sup_{f \in F} \left| \sum_{i=1}^{n} \sigma_i f(X_i) \right|$$

as follows:

$$P\{Y \neq f(X)\} \leq P_n\{Y \neq f(X)\} + \frac{R_n(F)}{2} + \sqrt{\frac{\ln(1/\delta)}{2n}},$$

where $P_n\{Y \neq f(X)\}$ is a training error and σ is a vector of Rademacher independent random variables and δ is a confidence interval. Our bound has similar inference to (Lanckriet *et al.*, 2004) and is stated for both versions of self-normalized topological kernels as follows:

$$P\{Y \neq f(X)\} \leq f_n\{Y \neq f(X)\} + \frac{1}{n\gamma} \mathrm{E}\left[\sum_{i=1}^{n} \sigma_i \cdot C(x_i, m) \right] + \sqrt{\frac{\ln(1/\delta)}{2n}}$$

$$C(x_i, m) = \sqrt{\sum_{j=1}^{m} \sup \left| \Omega_{down}\left(x_i^{(j)}\right) \right|^2 + m},$$

where m is input dimensionality of classification problem or analogously m=|τ| for the second expansion of topological kernel. Supremum in the square root of the helper term of the upper bound is inferred using Markov's inequality as follows:

$$\Omega_{down}\left(x^{(i)}\right) = \Pr\left[x^{(i)} \geq E[x]\right] \leq \frac{E\left[x^{(i)}\right]}{E[x]} = \sup \left| \Omega_{down}\left(x^{(i)}\right) \right|,$$

and reciprocal upper bound of topological measure for the up-regularity of *i-th* attribute is obtained following the reciprocal probability of:

$$\Omega_{up}\left(x^{(i)}\right) = \Pr\left[x^{(i)} < E[x]\right] = 1 - \Pr\left[x^{(i)} \geq E[x]\right].$$

Notice that we could get for only effective lower bound and surely this probability is upper-bounded by:

$$\Omega_{up}\left(x^{(i)}\right) = \Pr\left[x^{(i)} < E[x]\right] \leq 1 = \sup \left| \Omega_{up}\left(x_i^{(j)}\right) \right|.$$

Proof of the presented upper bound on generalization error of SVM using proposed topological kernel can be found in Appendix A.

4. EXPERIMENTS

In this section we present experimental setup and major results obtained from evaluating MKL method on RBF and proposed kernels.

4.1 Preliminaries

In our experiments we have tested proposed model under predefined C=10 (error tradeoff) value of soft-margin SVM that showed most comprehensible performance for imbalanced data sets (number of dimensions >> number of samples) and varying μ value of RBF Gaussian kernel that tradeoffs kernel smoothness and could be effectively estimated via simpleMKL framework (Rakotomamonjy et al., 2008). Mentioned above conditions were applied to all data sets just to ensure unbiased experimental results. Additionally for our primary data set we address two different versions of it:

1. Non-normalized dataset directly obtained by scanning microarray chip.
2. Normalized via OLIN (Futschik, 2005) and other techniques dataset.

Other important biomedical data sources were taken from UCI repository (Frank and Asuncion, 2010). Arrhythmia and Arcene data sets are assumed to be of particular interest because of high dimensionality but different (in comparison to microarrays) type of considered data origin. That's why

these data sets play essential role in benchmarking newly proposed topological expansion for linear and RBF kernels.

4.2 Experimental Setup

To verify and test topological and RBF kernels under selected performance measures it was decided to conduct following experimental setup that consists of some prefixed number of iterations (I=100 in our experimental setup) where every iteration verification set is composed of i.i.d. selected N samples (N=50 for all datasets) and all remaining samples are used as the training set (the only exception was made for Arcene data set for which verification set was randomly sampled from validation set). After fixing number of iterations each independent trial we have employed MKL approach to estimate optimal "tuning" parameters for SVM classifier. As it will be seen further we have conducted two separate sets of experiments for scaled and non-scaled invariants of evaluated datasets in order to estimate importance of scaling for proposed and RBF kernels. Finally each set of experiments was divided according to different origin of data sources (melanoma, gastric, prostate etc.) and every separate data source was evaluated jointly in all available invariants (normalized vs. non-normalized) on the same verification set (assuming independence of every iteration). Above mentioned scheme wasn't employed for UCI data sets. Because of initial segregation of Arcene training and validation data sets both sets reasonably weren't scaled and contributed to SVM "learning" in unrefined version. Arrhythmia data set was initially scaled and all the samplings were made from already preprocessed version of data source.

4.3 Results

In the following subsection we have summarized experimental results for all datasets under fixed C parameter and enclosed subspace for μ parameter with some initial guess of its corresponding scaling factor[1]. In current table we present performance measures obtained by MKL approach for initial SVM with standard RBF kernel and SVM with proposed topological kernel. Additionally we analyze and present results under scaled and non-scaled invariants of evaluated microarray datasets with provided number of selected by MKL kernels. All results are averaged across 100 independent trials and provided with standard deviations.

Table 1. Averaged performance measures

Dataset	Sensitivity	Specificity	Nr. of selected kernels	Error
Melanoma A[2] (RBF kernel)	0.962±0.037	0.814±0.075	1.980±0.140	0.108±0.038
Melanoma A (Topological kernel)	0.949±0.044	0.845±0.075	2.190±2.312	**0.101±0.040**
Melanoma B[3] (RBF kernel)	0.861±0.070	0.802±0.092	14.45±3.851	0.167±0.057
Melanoma B (Topological kernel)	0.978±0.037	0.874±0.067	5.540±8.407	**0.071±0.036**
Melanoma C[4] (RBF kernel)	0.972±0.034	0.791±0.089	1.950±0.219	0.113±0.046
Melanoma C (Topological kernel)	0.936±0.043	0.883±0.076	2.000±0.000	**0.088±0.043**
Melanoma D[5] (RBF kernel)	0.875±0.070	0.798±0.085	16.74±12.14	0.163±0.050
Melanoma D (Topological kernel)	0.984±0.024	0.881±0.066	5.760±9.342	**0.064±0.031**
Gastric A (RBF kernel)	0.903±0.070	0.247±0.097	15.98±5.596	0.401±0.069
Gastric A (Topological kernel)	0.917±0.065	0.241±0.090	19.36±2.772	**0.396±0.062**
Gastric B (RBF kernel)	0.726±0.110	0.505±0.142	5.810±4.948	0.381±0.070
Gastric B (Topological kernel)	0.839±0.078	0.517±0.120	1.960±0.197	**0.314±0.063**
Gastric C (RBF kernel)	0.821±0.089	0.317±0.110	8.910±9.151	0.416±0.076
Gastric C (Topological kernel)	0.843±0.088	0.360±0.151	14.97±9.066	**0.389±0.081**
Gastric D (RBF kernel)	0.719±0.114	0.533±0.130	12.39±12.31	0.376±0.068

[1] We have defined range of b_μ*10^[-10..10] with the step 0.25 resulting in a total of 81 kernels where b_μ is the corresponding scaling factor of μ stated as follows: $b_\mu = 1/2\sqrt{median(X)}$ where X is a vector of all dataset values

[2] Normalized scaled dataset
[3] Non-normalized scaled dataset
[4] Normalized non-scaled dataset
[5] Non-normalized non-scaled dataset

Gastric D (Topological kernel)	0.889±0.087	0.412±0.141	1.810±0.394	**0.338±0.077**
Prostate A (RBF kernel)	0.858±0.104	0.996±0.016	1.140±0.349	0.042±0.032
Prostate A (Topological kernel)	0.886±0.104	0.990±0.024	1±0	**0.039±0.032**
Prostate B (RBF kernel)	0.635±0.127	0.954±0.034	10.36±5.410	0.131±0.040
Prostate B (Topological kernel)	0.909±0.074	0.996±0.011	1.310±0.465	**0.028±0.020**
Prostate C (RBF kernel	0.868±0.116	0.931±0.050	1.060±0.239	0.088±0.043
Prostate C (Topological kernel)	0.608±0.145	1±0	1.080±0.273	0.102±0.048
Prostate D (RBF kernel)	0.664±0.143	0.959±0.032	9.380±10.02	0.118±0.047
Prostate D (Topological kernel)	0.901±0.087	0.999±0.006	1±0	**0.026±0.023**
Arrhythmia (RBF kernel)	0.674±0.065	0.817±0.062	5.980±8.713	0.252±0.034
Arrhythmia (Topological kernel)	0.684±0.056	0.829±0.050	2.450±0.642	**0.241±0.031**
Arcene (RBF kernel)	0.660±0.069	0.915±0.034	2±0	0.200±0.034
Arcene (Topological kernel)	0.683±0.075	0.917±0.037	2±0	**0.189±0.036**

4.4 Statistical Tests

In this section we present results of two-sided t-test applied to classification errors derived from 100 independent trials of MKL method for each invariant of evaluated dataset. The corresponding P-values that indicate level of confidence of the paired comparisons under null-hypothesis of equal error means are represented in Table 2 and Table 3. Note that we have conducted additionally two joint comparisons for normalized and non-normalized invariants of microarray datasets (assuming identical validation sets) in order to show that topological kernel performs significantly (in terms of achieved P-values) better than RBF kernel especially on non-normalized data.

Table 2. P-values that indicate the confidence level of significance when comparing proposed topological kernel to RBF kernel in terms of classification error

Dataset	Topological kernel vs. RBF kernel
Melanoma A	0.17626
Melanoma B	**0**
Melanoma C	**0.00014896**
Melanoma D	**0**
Gastric A	0.59833
Gastric B	**1.8182e-11**
Gastric C	**0.017626**
Gastric D	**0.00023403**
Prostate A	0.48182
Prostate B	**0**
Prostate C	**0.026427**
Prostate D	**0**
Arrhythmia	**0.026**
Arcene	**0.025**

Table 3. P-values that indicate the confidence level of significance when comparing normalized to non-normalized datasets in terms of classification error (assuming strictly topological kernel to RBF kernel comparison)

Dataset	Normalized vs. non-normalized	Non-normalized vs. normalized
Melanoma AL	**1.8181e-11**	**0**
Melanoma BL	**6.6613e-16**	**0**
Gastric AL	**0**	0.11797
Gastric BL	**1.0951e-11**	0.22854
Prostate AL	**0.00016227**	**0**
Prostate BL	**0**	**0.020899**

4.5 Analysis

As we could see from performance measures on different data sources almost everywhere topological kernel usage attains the same or even better results bringing some useful discrimination capabilities to SVM classifier. The achieved result could be easily verified on two benchmark data sets: Arcene and Arrhythmia,

where topological kernel outperforms RBF kernel in terms of accuracy and classification rate. For normalized microarray datasets[6] newly proposed kernel attains just a slight improvement in total generalization error but for non-normalized one it surprisingly leads to a major classification performance boost that seems like a very promising result because it outperforms even normalized datasets using the same newly proposed kernel. The significance of this result could be clearly proven by P-values attained from comparison of classification errors on different invariants of microarray datasets. For instance in comparison with basis RBF kernel for non-normalized melanoma dataset our kernel improves accuracy almost by 10%. It brings us to assumption that even very slight (and might be improper) normalization can decrease classifier discrimination capabilities and classification accuracy. Another interesting assumption is related to more precise (in estimating allocation in Hilbert space) nature of topological measures being incorporated into proposed kernel and their superior capabilities in estimating similarity by the use of dot product in highly dimensional input spaces.

5. CONCLUSION

In this paper we propose new topological expansion for smoothing the kernels used for classification tasks in biomedicine and molecular biology. Attained results show that proposed kernel outperforms SVM with basis RBF kernel and works even better on non-normalized[6] highly dimensional microarray data. This result opens new perspectives for usage of raw highly through-output microarray technologies in practical biomedicine without proper pre-processing and normalization of collected data. In this paper we present upper bound on SVM generalization error with proposed kernel and show that this bound depends on input dimensionality and "signal" within each sample as a whole and expected values of its attributes as well. Finally we examined two additional highly dimensional biomedical datasets (Arcene and Arrhythmia) from UCI repository and conclude that superiority of proposed kernel is confidently feasible not only on microarray-related diagnostics but also on other very important biomedical problems.

APPENDIX A

In this appendix we prove presented in Section 3.2 generalization error bound for self-normalized version of proposed kernel. Firstly rewrite original bound presented by (Bartlett and Mendelson, 2002):

$$P\{Y \neq f(X)\} \leq P_n\{Y \neq f(X)\} + \frac{R_n(F)}{2} + \sqrt{\frac{\ln(1/\delta)}{2n}}$$

Now infer Rademacher complexity with respect to general kernel mapping as follows:

$$\begin{aligned} R_n(F) &= \frac{2}{n}\sup_{f\in F}\left|\sum_{i=1}^{n}\sigma_i f(X_i)\right| \\ &= \frac{2}{n}\max_{\|w\|\leq 1/\gamma}\left|\sum_{i=1}^{n}\sigma_i\langle w, \Phi(X_i)\rangle\right| \\ &\leq \frac{2}{n}\max_{\|w\|\leq 1/\gamma}\|w\|\cdot\sum_{i=1}^{n}\sigma_i\|\Phi(X_i)\| \\ &= \frac{2}{n\gamma}\sum_{i=1}^{n}\sigma_i\|\Phi(X_i)\|. \end{aligned}$$

Now reformulate term $\|\Phi(X_i)\|$ by the means of newly proposed kernel, upper bound on topological measures and separation of input dimensionality for down- and up-regularity measures:

[6] Meaning global cross-sample normalization

$$\|\Phi(X_i)\| = \sqrt{\langle \Phi(X_i), \Phi(X_i) \rangle}$$
$$\leq \sqrt{\sum_{j=1}^{m} \left[\sup \left| \Omega_{down}\left(x_i^{(j)}\right) \right|^2 + \sup \left| \Omega_{up}\left(x_i^{(j)}\right) \right|^2 \right]}$$
$$= \sqrt{\sum_{j=1}^{m} \left[\frac{E\left[x_i^{(j)}\right]^2}{E\left[x_i\right]^2} + 1^2 \right]}$$
$$= \sqrt{\sum_{j=1}^{m} \sup \left| \Omega_{down}\left(x_i^{(j)}\right) \right|^2 + m}.$$

By inserting the last term of equation into Rademacher complexity and corresponding bound yielded by (Bartlett and Mendelson, 2002) we have proved presented in this paper upper bound on SVM generalization error using proposed topological kernel.

REFERENCES

Rakotomamonjy, A. *et al.*, 2008. SimpleMKL. *Journal of Machine Learning Research*, Vol. 9, pp 2491-2521.

Kalnina, Z. *et al.*, 2008a. Autoantibody profiles as biomarkers for response to therapy and early detection of cancer. *Current Cancer Therapy Reviews*, Vol. 4 (2), pp 149-156.

Kalnina, Z. *et al.*, 2008b. Evaluation of T7 and Lambda phage display systems for survey of autoantibody profiles in cancer patients. *J Immunol Methods*, Vol. 334(1-2), pp 37-50.

Vapnik, V., 1995. *The Nature of Statistical Learning Theory*. Springer-Verlag.

Lanckriet, G. *et al.*, 2004. Learning the Kernel Matrix with Semidefnite Programming. *Journal of Machine Learning Research*, Vol. 5, pp 27-72.

Chapelle, O. *et al.*, 2002. Choosing Multiple Parameters for Support Vector Machines. *Machine Learning*, Vol. 46, pp 131–159.

Bartlett, P. L. and Mendelson, S., 2002. Rademacher and Gaussian complexities: Risk bounds and structural results. *Journal of Machine Learning Research*, Vol. 3, pp 463-482.

Sonnenburg, S. *et al*., 2006. Large scale multiple kernel learning. *Journal of Machine Learning Research*, Vol. 7(1), pp 1531–1565.

Bach, F. *et al.*, 2004. Multiple kernel learning, conic duality, and the SMO algorithm. *In Proceedings of the 21st International Conference on Machine Learning*. Montreal, Canada, pp. 41–48.

Futschik, M., 2005. *Introduction to OLIN package* [ONLINE], Available at: http://www.bioconductor.org/packages/2.3/bioc/vignettes/OLIN/inst/doc/OLIN.pdf [Accessed 25 March 2010].

Frank, A. and Asuncion, A., 2010. *UCI Machine Learning Repository*, Available at: http://archive.ics.uci.edu/ml [Accessed 25 March 2010] Irvine, CA: University of California, School of Information and Computer Science.