Regularized Knowledge-Based Kernel Machine

Olutayo O. Oladunni and Theodore B. Trafalis

School of Industrial Engineering, The University of Oklahoma
202 West Boyd, CEC 124 Norman, OK 73019 USA
{tayo, ttrafalis}@ou.edu

Abstract. This paper presents a knowledge-based kernel classification model for binary classification of sets or objects with prior knowledge. The prior knowledge is in the form of multiple polyhedral sets belonging to one or two classes, and it is introduced as additional constraints into a regularized knowledge-based optimization problem. The resulting formulation leads to a least squares problem that can be solved using matrix or iterative methods. To evaluate the model, the experimental laminar & turbulent flow data and the Reynolds number equation used as prior knowledge were used to train and test the proposed model.

1 Introduction

In data mining applications, incorporation of prior knowledge is usually not considered because most algorithms do not have the adequate means for incorporating explicitly such types of constraints. In the case of Support Vector Machines (SVMs) model [1], Fung et al. [2], [3] developed explicit formulations for incorporating prior knowledge in the form of multiple polyhedral sets belonging to one or more categories into a linear and nonlinear classification model.

The main motivation of this paper is to develop a nonlinear kernel approach based on the Tikhonov regularization scheme for knowledge-based classification discrimination. The proposed model problem is the kernel-based formulation of the multi-classification model in [4] for a two-class classification problem. The feature of the proposed model is that it leads to a least squares problem that can be solved by solving a linear system of equations that reduces computational time. In contrast, Fung et al. [2], [3] solve a linear programming optimization problem.

In this work, a combination of the prior knowledge sets together with the concept of the least squares SVM (LS-SVM) [5] and regularization based LS-SVM [6] results into a linear system of equations. Our approach is different in the sense that, the proposed minimization problem is an unconstrained optimization problem resulting into a smaller linear system of equations than those proposed in [5], [6]. By expressing the normal vector w as a linear combination of the data points, it turns out that in the least squares knowledge based formulation the prior knowledge(s) for nonlinear classification is the precise implication of the prior knowledge(s) for linear classification. This differs from the knowledge based model by Fung et al. [3], where the prior knowledge(s) for nonlinear classification is not the precise implication of the prior knowledge(s) for linear classification. In their formulation, the knowledge constraints are

Y. Shi et al. (Eds.): ICCS 2007, Part I, LNCS 4487, pp. 176–183, 2007.

kernelized so as to fit a standard [7] linear programming formulation for nonlinear kernel classification.

The regularized knowledge-based kernel classification model can be considered as a least squares knowledge based nonlinear kernel formulation of Fung et al. [3]. Note that the resulting regularized knowledge-based model problem minimizes the classical regularization objective function with strict L_2 norm functions. This norm guarantees the differentiability of the objective function, and as result a linear system of equations is derived. In the knowledge based model by Fung et al. [3], the differentiability property is lost because in its model problem the functions are L_1 norm functions, which further restrain the solution of the model problem and also requires an LP solver to obtain a solution.

Benefits of the regularized knowledge-based kernel classification model includes: the reduction of a classification problem to a smaller linear system of equation, the ability to provide fast solutions which require no special solvers, the ability to provide explicit solution in terms of the given data and prior knowledge sets, and the ability to provide effective and robust classifiers as a result of the bounding planes within a cluster of points (margin increase).

2 Prior Knowledge in Two-Class Classification

Suppose that in addition to the points belonging to a class, there is prior information belonging to one or two categories. The knowledge sets [2] in an n dimensional space are given in the form of a polyhedral set determined by the set of linear equalities and linear inequalities. The polyhedral knowledge set $\{x \in R^n \mid Bx \le b\}$ or $\{x \in R^n \mid \overline{B}x = \overline{b}\}$ should lie in the halfspace $\{x \in R^n \mid x^T w \ge \gamma + 1\}$ where $B \in R^{g_u \times n}$, $\overline{B} \in R^{g_{\overline{u}} \times n}$, $b \in R^{g_u}$ and $\overline{b} \in R^{g_{\overline{u}}}$ are the prior information belonging to class 1. g_u or $g_{\overline{u}}$ is the number of prior knowledge (equality or inequality) constraints in class 1. d_v or $d_{\overline{v}}$ is the number of prior knowledge (equality or inequality) constraints in class 2. Therefore, the following implications must hold for a given (w, γ), where w is a normal vector and γ is the location of the optimal separating plane relative to the origin:

$$\left\{Bx \le b \Rightarrow x^T w \ge \gamma + 1\right\} \text{ or } \left\{\overline{B}x = \overline{b} \Rightarrow x^T w \ge \gamma + 1\right\}. \tag{2.1}$$

Using the nonhomogeneous Farkas theorem of the alternative [8] or, using duality in linear programming (LP) [9], equations (2.1) can be transformed into a set of knowledge constraints [2].

$$\left\{B^T u + w = 0,\ b^T u + \gamma + 1 \le 0,\ u \ge 0\right\} \text{ or } \left\{\overline{B}^T \overline{u} + w = 0,\ \overline{b}^T \overline{u} + \gamma + 1 \le 0\right\}. \tag{2.2}$$

To formulate the nonlinear counterpart of the linear classification, the primal variable w is replaced by its equivalent dual representation $w = \overline{A}^T Y \alpha$, where α is the vector of dual variables, $\overline{A} = [A^{(1)T}\ A^{(2)T}]^T$ whose rows are the points belonging to classes 1 and 2, and Y is a diagonal matrix whose diagonals are +1 for points in class 1 and -1 for points in class 2.

Let $A \in R^{m \times n}$ *and* $B \in R^{n \times k}$. *The* ***kernel*** $K(A,B)$ *maps* $R^{m \times n} \times R^{n \times k}$ *into* $R^{m \times k}$. The behavior of a kernel matrix strongly relies on Mercer's condition [10], [11] for symmetric kernel functions, i.e., a kernel matrix is a positive semidefinite (PSD) matrix.

Using the dual representation of w and applying the kernel definition, a more complex classifier can be determined, and the nonlinear optimal separating hyperplane is given as:

$$K(x^T, \overline{A}^T)Y\alpha = \gamma . \tag{2.3}$$

The implications for a given $(A,\ Y,\ \alpha,\ \gamma)$ now becomes the following [3]:

$$\left\{ Bx \le b \Rightarrow x^T \overline{A}^T Y\alpha \ge \gamma + 1 \right\} \text{ or } \left\{ \overline{B}x = \overline{b} \Rightarrow x^T \overline{A}^T Y\alpha \ge \gamma + 1 \right\}, \tag{2.4}$$

and the prior knowledge constraints can be rewritten as:

$$\left\{ B^T u + \overline{A}^T Y\alpha = 0,\ \ b^T u + \gamma + 1 \le 0,\ \ u \ge 0 \right\} \text{ or } \left\{ \overline{B}^T \overline{u} + \overline{A}^T Y\alpha = 0,\ \ \overline{b}^T \overline{u} + \gamma + 1 \le 0 \right\}. \tag{2.5}$$

Notice that there is no kernel present in the prior knowledge, and as a result equation (2.4) is the precise implications of equations (2.1). In the subsequent section, when the prior knowledge constraint is incorporated into a classification model, the kernel will be employed to obtain nonlinear classifiers.

3 Regularized Knowledge-Based Kernel Classification Machine

Consider a problem of classifying data sets with prior knowledge in R^n that are represented by a data matrix $A^{(i)} \in R^{m_i \times n}$, where $i = 1, 2$, and knowledge sets $\{x \mid B^{(1)}x \le b^{(1)}\}$ or $\{x \mid \overline{B}^{(1)}x = \overline{b}^{(1)}\}$ belonging to class 1, $\{x \mid B^{(2)}x \le b^{(2)}\}$ or $\{x \mid \overline{B}^{(2)}x = \overline{b}^{(2)}\}$ belonging to class 2. Let $A^{(1)}$ be a $m_1 \times n$ matrix whose rows are points in class 1, and m_1 is the number of data in class 1. Let $A^{(2)}$ be a $m_2 \times n$ matrix whose rows are points in class 2, and m_2 is the number of data in class 2. This problem can be modeled through the following optimization problem:

$$\min_{\alpha,\gamma,a,c} f_{NT_R KKM}(\alpha,\gamma,a,c) = \begin{bmatrix} \frac{\lambda}{2}\|\alpha\|^2 + \frac{1}{2}\|YKY\alpha - \gamma Ye - e\|^2 + \\ \frac{1}{2}\left[\left\|B_u^T a + I_u \overline{A}^T Y\alpha\right\|^2 + \left\|B_v^T c - I_v \overline{A}^T Y\alpha\right\|^2\right] + \\ \frac{1}{2}\left[\left\|B_{bu}^T a + \gamma e_u + e_u\right\|^2 + \left\|B_{bv}^T c - \gamma e_v + e_v\right\|^2\right] \end{bmatrix}. \tag{3.1}$$

Where α is the vector of dual variables; γ is the location of the optimal separating plane; λ is a regularization parameter; $a = [u^T, \overline{u}^T]^T$ is a vector of all multipliers referring to class 1, and $c = [v^T, \overline{v}^T]^T$ is a vector of all multipliers referring to class 2. $K = K(\overline{A}, \overline{A}^T)$ is a kernel matrix; matrix $\overline{A} = [A^{(1)T}\ \ A^{(2)T}]^T$ whose rows are the points belonging to classes 1 & 2 points respectively; Y is a diagonal matrix whose diagonals are +1 for points in class 1 and -1 for points in class 2; and vector $e = [e^{(1)T}\ \ e^{(2)T}]^T$

is a vector of ones. Matrices $B_u \& B_v$ are diagonal block matrices whose diagonals contain knowledge sets belonging to class 1 & 2. The diagonals of B_u are $B^{(1)} \in R^{g_u \times n}$, $\bar{B}^{(1)} \in R^{g_{\bar{u}} \times n}$ and of B_v are $B^{(2)} \in R^{d_v \times n}$, $\bar{B}^{(2)} \in R^{d_{\bar{v}} \times n}$. Matrices $B_{bu} \& B_{bv}$ are created where the diagonals of B_{bu} are $b^{(1)} \in R^{g_u \times 1}$, $\bar{b}^{(1)} \in R^{g_{\bar{u}} \times 1}$ and of B_{bv} are $b^{(2)} \in R^{d_v \times 1}$, $\bar{b}^{(2)} \in R^{d_{\bar{v}} \times 1}$. Matrices $I_u = [I_{(u)}^T \ \bar{I}_{(u)}^T]^T$ & $I_v = [I_{(v)}^T \ \bar{I}_{(v)}^T]^T$ are block matrices, where $I_{(u)}$, $\bar{I}_{(u)}$, $I_{(v)}$, $\bar{I}_{(v)}$ $\in R^{n \times n}$ are identity matrices. Vectors e_u & e_v consist of entries $e_1, \bar{e}_1 \in R$ & $e_2, \bar{e}_2 \in R$, where $e_1, \bar{e}_1, e_2, \bar{e}_2$ are equal to one. Vectors e_u, e_v are vectors of ones, where each entry corresponds to a vector $b^{(1)}, \bar{b}^{(1)}$ ($b^{(2)}, \bar{b}^{(2)}$) respectively.

The (α, γ) taken from a solution of (3.1) generates the nonlinear separating surface (2.3). Problem (3.1) was formulated using the concept of penalty functions [9], and it is called the nonlinear Tikhonov regularization [12] knowledge-based kernel machine classification model (NT_RKKM). It is a binary classification formulation of the multi-classification model in [4] to accommodate nonlinearly separable patterns with knowledge sets, and explicit solutions in dual space in terms of the given data can be derived. Problem (3.1) is slightly different from the knowledge based model by Fung et al. [4], which is a linear programming formulation. The difference lies in the selection of the norm distance and the squared error with regularization term (least squares).

Below is the explicit solution to NT_RKKM in terms of the given data:

$$\gamma = \bar{h}(d\alpha - t). \tag{3.2}$$

$$\bar{M}\alpha = \bar{z} \Rightarrow \alpha = \bar{M}^{-1}\bar{z}, \text{ where } \alpha \in R^m. \tag{3.3}$$

Note that $\bar{h}$, d, t, $\bar{M}$, $\bar{z}$ are defined as follows.

$$\bar{h} = \left[e^T YYe + e_u^T\left[e_u - B_{bu}^T U B_{bu} e_u\right] + e_v^T\left[e_v - B_{bv}^T V B_{bv} e_v\right]\right]^{-1}. \tag{3.4}$$

$$\begin{aligned} d &= \left[e^T Y\bar{K} + e_u^T B_{bu}^T U K_{Bu}^T Y + e_v^T B_{bv}^T V K_{Bv}^T Y\right] \\ t &= \left[e^T Ye + e_u^T\left[e_u - B_{bu}^T U B_{bu} e_u\right] - e_v^T\left[e_v - B_{bv}^T V B_{bv} e_v\right]\right] \end{aligned}. \tag{3.5}$$

$$\begin{aligned} &U = \left(B_u B_u^T + B_{bu} B_{bu}^T\right)^{-1},\ V = \left(B_v B_v^T + B_{bv} B_{bv}^T\right)^{-1},\ \bar{K} = YKY \\ &K_u = K(\bar{A}I_u^T, I_u\bar{A}^T),\ K_v = K(\bar{A}I_v^T, I_v\bar{A}^T),\ K_{Bu} = K(\bar{A}I_u^T, B_u^T),\ K_{Bv} = K(\bar{A}I_v^T, B_v^T) \end{aligned}. \tag{3.6}$$

$$\bar{M} = \begin{bmatrix} \lambda I + \bar{K}\left[\bar{K} - Y\bar{h}ed\right] + \\ Y\left[K_u Y - K_{Bu} U\left[K_{Bu}^T Y + B_{bu}\bar{h}e_u d\right]\right] + \\ Y\left[K_v Y - K_{Bv} V\left[K_{Bv}^T Y + B_{bv}\bar{h}e_v d\right]\right] \end{bmatrix},\ \bar{z} = \begin{bmatrix} \bar{K}^T\left[e - Y\bar{h}et\right] + \\ YK_{Bu} U B_{bu}\left[e_u - \bar{h}e_u t\right] - \\ YK_{Bv} V B_{bv}\left[e_v + \bar{h}e_v t\right] \end{bmatrix}. \tag{3.7}$$

Minimizing the L_2 norm of α guarantees a solution for a positive tradeoff constant. It is evident from matrix $\bar{M}$ that only the diagonals change with any change in the trade-off constant, implying we can always get a diagonally dominant matrix $\bar{M}$ which can ensure a solution. However, if the first term in $\bar{M}$ is replaced with a kernel matrix, then for any constant, all elements in matrix $\bar{M}$ also increase. Therefore, we cannot guarantee a diagonally dominant matrix $\bar{M}$ and hence we cannot always guarantee a solution when minimizing the square norm of the linear combination of w in dual space.

Assume that U, V and $\bar{M}$ are invertible matrices. Solution (3.3) provides an estimate of the dual variables. Such a linear system is easier to solve than a quadratic or linear programming formulation. Methods for solving the system include matrix decomposition methods, or iterative based methods [9], [13]. Its solution involves the inversion of an $m \times m$ dimensional matrix.

The decision function for classifying a point x is given by:

$$D(x) = sign\left[K(x^T, \bar{A}^T) Y\alpha - \gamma \right] = \begin{cases} +1, & \text{if point } (x) \text{ is in class } A^{(1)} \\ -1, & \text{if point } (x) \text{ is in class } A^{(2)} \end{cases} . \tag{3.8}$$

4 Laminar/Turbulent Flow Pattern Data Set

The fluid flow data uses flow rate, density, viscosity, borehole diameter, drill collar diameter and mud type to delineate the flow pattern (laminar, +1; turbulent, -1) of the model. There are 92 data-points and 5 attributes, 46 instances for laminar flow pattern and 46 instances turbulent flow pattern [14], [15], [16]. The attributes are as follows: ρ, density of fluid (lbm/gal – continuous variable); q, flow rate (gal/min – continuous variable); (d_2+d_1), summation of borehole and drill collar (OD) diameter (in – continuous variable); μ_p, plastic viscosity (cp – continuous variable); Mud type, water based mud (1) oil based mud (2) (categorical variable).

Prior Knowledge: In addition to the fluid flow data, the Reynold's equation [17] for a Bingham plastic model is used as prior knowledge to develop a knowledge based classification model. As additional constraints, the prior knowledge will represent the transition equations which delineates laminar from turbulent flows. Since the flow pattern data are scaled by taking the natural logarithm of each instance, the prior knowledge needs to be scaled by also considering a natural logarithm transformation of the equations. Below is the equivalent logarithmic transformation for Reynold's equation [17]:

$$\Rightarrow \begin{cases} \ln(\rho) + \ln(q) - \ln(d_2 + d_1) - \ln(\mu_p) \leq 1.9156 - \varepsilon \rightarrow x^T w - \gamma \geq +1 \text{ (Laminar)} \\ \ln(\rho) + \ln(q) - \ln(d_2 + d_1) - \ln(\mu_p) \geq 1.9156 + \varepsilon \rightarrow x^T w - \gamma \leq -1 \text{ (Turbulent)} \end{cases}, \tag{4.1}$$

where ε is a deviation factor, a small fraction perturbing the critical Reynold's number. The dataset where divided up as follows; 50% of the whole data is used as training data, while the remaining 50% was used as testing data.

5 Computational Results

In this section, the results of the analyzed data sets and prior sets described in section 4, are presented and discussed. The NT_RKKM model is used to train the data sets with prior knowledge. To demonstrate the uniqueness of the formulations, a comparison between the models above was conducted. The comparisons were made by evaluating a performance parameter (misclassification error) defined below:

$$\beta = 1 - \left(\frac{Total\ number\ of\ correctly\ classified\ points}{Total\ number\ of\ observed\ points} \right). \qquad (5.1)$$

β represents the overall misclassification error rate, i.e., the fraction of misclassified points for a given data set. For 100% classification, $\beta = 0$. The tradeoff constant considered is within the interval 0 – 100, and the deviation factor $\varepsilon = 0.01$. Further comparisons were made between the NT_RKKM model, LT_RKSVM [16] model problem, the traditional SVM [1], [14], and the mixed integer programming kernel based classifiers ($MIPKC_{ps}$ & $MIPKC_p$, [15]). The polynomial kernel, $k(x_i, x) = (x_i^T x + 1)^p$, where p is the degree of the polynomial ($p = 2$) was used.

Results of the fluid flow pattern data with prior knowledge information trained on the NT_RKKM model, and compared with the LT_RKSVM, SVM, $MIPKC_{ps}$, & $MIPKC_p$ model, are shown in Tables 1 & 2. It should be noted that β (error rate) in the Tables are defined by (5.1), and the computing (cpu) time is measured in seconds. All computations and experiments were performed using MATLAB [18] for NT_RKKM, LT_RKSVM & SVM model [19], and CPLEX OPL [20] for the mixed integer programming kernel based models.

Table 1. Sample and Average random sample validation test error rate for NT_RKKM on Fluid Flow Pattern Data (varying tradeoff, λ)

Statistics	Tradeoff	Test 1	Test 2	Test 3	Test 4	Average
Error	0	-	-	-	-	-
Cpu time		-	-	-	-	-
Error	1	0.0000	0.0278	0.0000	0.0000	0.0070
Cpu time		*0.0200*	*0.0100*	*0.0200*	*0.0100*	*0.0150*
Error	5	0.0000	0.0000	0.0000	0.0000	**0.0000**
Cpu time		*0.0400*	*0.0100*	*0.0100*	*0.0200*	*0.0200*
Error	10	0.0000	0.0000	0.0000	0.0000	**0.0000**
Cpu time		*0.0200*	*0.0800*	*0.0200*	*0.0100*	*0.0325*
Error	50	0.0000	0.0000	0.0000	0.0278	0.0070
Cpu time		*0.0100*	*0.0100*	*0.1310*	*0.0700*	*0.0553*
Error	100	0.0000	0.0000	0.0556	0.0278	0.0209
Cpu time		*0.0200*	*0.0100*	*0.0100*	*0.0200*	*0.0150*

Table 2. Average error, accuracy & cpu. time of NT_RKKM, LT_RKSVM, MIPKC & SVM models

Average	**NT_RKKM**	**LT_RKSVM**	**MIPKCps**	**MIPKCp**	**SVM**	**SVM [14]**
Error	0.0000	0.0139	0.0139	0.0348	0.0209	0.0500
Accuracy	1.0000	0.9861	0.9861	0.9652	0.9792	0.9500
Cpu time	0.0200	0.0100	49.2975	83.9400	0.4250	0.4600

Tables 1 & 2 contain results for the NT_RKKM, LT_RKSVM, $MIPKC_{ps}$, $MIPKC_p$, & SVM model on the fluid flow pattern classification data. The models in comparison generally report promising error rates. The NT_RKKM model reports the smallest error (0). This means that the testing data set was correctly classified (100% classification).

The error rate (0.0139) for the model with prior knowledge (LT_RKSVM) performs in the same capacity as the data driven mixed integer programming model ($MIPKC_{ps}$). The computation time of both the NT_RKKM and LT_RKSVM model are comparable, and both clearly outperform the other learning models. The SVM models reports the next best computation time and the MIPKC model reports the worst time.

The fast solution of the LT_RKSVM model is due the formulation of the problem that is analyzed analytically in the primal (input) space, i.e., the dimension n of the classification problem is equal to the number of attributes (variables) of the data set; in our case n = 5. The fast solution of the NT_RKKM model is due to the small number of data in the test set (m = 46 test observations) and when possible, the avoidance of iterative methods in computing its solution. The next best computation time to the NT_RKKM model is the one of the SVM model. Its fast solution is also due to the small number of test observations (m = 46 test observations). Note that in this case iterative methods are employed to find its solution. The $MIPKC_{ps}$, $MIPKC_p$ models require more time because their formulation is based on integer and mixed integer programming techniques which generally perform more computations in obtaining a solution than its linear counterparts such as SVM, LT_RKSVM and NT_RKKM.

6 Conclusion

In this paper, a binary classification model called the nonlinear classification Tikhonov regularization knowledge-based kernel machine (NT_RKKM) is described. This model problem is an unconstrained optimization problem for discriminating between two disjoint sets. The proposed model can be considered as a least squares formulation of Fung et al. [3] knowledge-based nonlinear kernel model. The model was applied to the laminar/turbulent fluid flow data. Comparisons were made with the linear and nonlinear counterpart formulations, and best statistics were obtained by performing training on the NT_RKKM model. The NT_RKKM model can be applied to determine the flow patterns of fluid flow with different rheology. The fluid model in this paper is a non-Newtonian fluid, pseudoplastic model (Bingham plastic) with two flow patterns (laminar & turbulent flows). The same concept can be applied to a Newtonian or non-Newtonian fluid with three flow patterns (laminar, transition and turbulent).

References

1. Vapnik, V.: The Nature of Statistical Learning Theory. New-York: Springer-Verlag (1995)
2. Fung, G., Mangasarian, O.L., Shavlik, J.W.: Knowledge-based Support Vector Machine Classifiers. Neural Information Processing Systems 2002 (NIPS 2002), Vancouver, BC, December 10-12, (2002)
3. Fung, G., Mangasarian, O.L., Shavlik, J.W.: Knowledge-based Nonlinear Kernel Classifiers, in Manfred Warmuth and Bernhard Scholkopf (eds), Proceedings Conference on Learning Theory (COLT/Kernel 03) and Workshop on Kernel Machines, Washington, D.C., August 24 - 27, (2003) 102-113

4. Oladunni, O.O., Trafalis, T.B., Papavassiliou, D.V.: Knowledge-based multiclass support vector machines applied to vertical two-phase flow, ICCS 2006, Lecture notes in Computer Science, (V.N. Alexandrov et al., eds.), Part I, LNCS 3991. Springer-Verlag, Berlin Heidelberg (2006) 188-195
5. Suykens, J.A.K. and Vandewalle, J.: Least Squares Support Vector Machine Classifiers. Neural Processing Letters, 9, (1999b) 293-300
6. Pelckmans, K., Suykens, J.A.K., De Moor, B.: Morozov, Ivanov and Tikhonov Regularization based LS-SVMs. In Proceedings of the International Conference On Neural Information Processing (ICONIP 2004), Calcutta, India, Nov. 22-25, (2004)
7. Mangasarian, O.L.: Generalized Support Vector Machines. Advances in Large Margin Classifiers, (A. Smola, P. Bartlett, B. Schölkopf, and D. Schuurmans, eds.), MIT Press, Cambridge, MA (2000) ftp://ftp.cs.wisc.edu/pub/dmi/tech-reports/99-14.ps
8. Mangasarian, O.L.: Nonlinear Programming. Philadelphia, PA: SIAM (1994)
9. Bazaraa, M.S., Sherali, H.D., Shetty, C.M.: Nonlinear Programming – Theory and Algorithms. John Wiley & Sons, Inc. (1993)
10. Burges, C.J.C.: A Tutorial on Support Vector Machines for Pattern Classification. Data Mining and Knowledge Discovery, 2(2): 121-167, (1998)
11. Cristianini, N., Shawe-Taylor, J.: Support Vector Machines and other Kernel-based Learning Methods; Cambridge University Press, (2000)
12. Tikhonov, A.N., Arsenin, V.Y.: Solution of Ill-Posed Problems. Winston, Washington D.C., (1977)
13. Lewis, J. M., Lakshmivarahan, S., Dhall, S.: Dynamic Data Assimilation. Cambridge University Press, (2006)
14. Trafalis, T.B.; Oladunni, O.: Single Phase Fluid Flow Classification via Neural Networks & Support Vector Machine. Intelligent Engineering Systems Through Artificial Neural Networks, (C.H. Dagli, A.L. Buczak, D. L. Enke, M.J. Embrechts, and O. Ersoy, eds.), ASME Press, 14, (2004) 427-432
15. Oladunni, O.; Trafalis, T.B.: Mixed-Integer Programming Kernel-Based Classification, WSEAS Transactions on Computers, Issue 7, Vol. 4: 671-678, July (2005)
16. Oladunni, O.; Trafalis, T.B.: Single Phase Laminar & Turbulent Flow Classification in Annulus via a Knowledge-Based Linear Model. Intelligent Engineering Systems Through Artificial Neural Networks, (C.H. Dagli, A.L. Buczak, D. L. Enke, M.J. Embrechts, and O. Ersoy, eds.), ASME Press, 16, (2006) 599-604
17. Bourgoyne, A.T. Jr., Chenevert, M.E., Millheim, K.K., Young, F.S. Jr.: Applied Drilling Engineering, SPE Textbook Series Vol. 2, Richardson, TX, p. 155, (1991)
18. MATLAB User's Guide.: The Math-Works, Inc., Natwick, MA 01760, 1994-2003. http://www.mathworks.com
19. Gunn, S.T.: Support Vector Machine for Classification and Regression, Technical Report, Department of Electronic and Computer Science, University of Southampton, (1998)
20. ILOG OPL STUDIO 3.7.: Language Manual, ILOG, S.A. Gentily, France and ILOG, Inc., Mountain View California, USA. http://www.ilog.com/products/oplstudio/, (2003)