ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)
Vol 5, Issue 2, February 2018

Indeed A Big Technology: The Kernel methods

[1] Sandhya Pundhir, [2] Varsha Kumari, [3] M. Q. Rafiq
[1] KIT, Pitampura, Delhi, [2] GLA University, Mathura, [3] JPU, Anoopshahar

***Abstract* - Presently there has been good interest in computing similarity for data mining and machine learning particularly. In this paper, we are discussing Kernel methods. Kernel methods are used pervasively mainly because of its large number of applications and scientific challenges. It has the capability to model real-world data and give efficient solutions to real-world problems. Such solutions given are one of its kind the most accurate and efficient as compared to the other existing ones. Along with the popular applications of kernel this paper mainly gives a basic understanding of fundamental aspects of kernel methods their underlying principles and formulas used. Various aspects of kernel approach are discussed. Some simulation results are shown in the discussed kernel methods and compared with new proposed kernel design.**

I. INTRODUCTION

Mostly real world data has no readily available representation as a single table or structured data. But present data mining methods mostly focused on single table. So to apply traditional data mining methods an extensive pre processing has to be performed. By that the computational cost and processing time has become a road block to effective data mining.

Hence all focus goes on a method which has less computational cost as well as easy to understand. Also results produced by the method should be interpretable. So here comes kernel trick or kernel method. A simple way of describing a kernel is a similarity function which is used in machine learning algorithm. Kernel may also be known as another name for the covariance function it takes set of input and find out how similar they are. As under some conditions every kernel function can be expressed as dot product in a feature space (mercer theorem) and many machine learning algorithms can be expressed in terms of dot products. The kernel method may be used for the different type of variables such as continuous variable, binary variable, nominal variable as well as ordinal variable.

Kernel methods have less computational cost because they can be applied in high dimensional feature spaces without having the high cost of computing to map data. They can easily do modeling of highly nonlinear data. Kernel methods have shown number of attractive features with empirical excellent performance. One of the strength of kernel methods is that they model non parametric modeling technique where it in not required to specify beforehand number of basic functions.

Kernel method algorithms are like the classical approaches with theoretical elegance that have closed form analytic solution with flexibility offered by modern nonlinear approaches. These methods have natural ability to deal with data that is not linearly separable and does not require user specific regularization parameter for penalizing misclassification. Kernel method is capable of handling very large datasets with modest computational load.

One of the critical parts of the kernel based learning algorithm is the choice of a best kernel method and the optimal parameters according to the data used. Kernel increases flexibility by increasing allowed similarity measures and make it possible to work with non vector form data [7][13]. Kernel methods when used with neural network leads to computational intelligence. Several international organizations, conferences and journals are dealing closely with this concept.

Overall the purpose of this paper is to show kernel trick, current scope from its past and present applications and uses while highlighting the answers of following:
•To illustrate basics of kernel methods so that one can understand them and apply as well.
•To know the computations achieved from usage of kernel trick.
•To know the evaluation and measurement methods such as standard or the popularly used ones.
•Study of simulation results of various existing kernel methods and a new proposed design.

This paper is organized as follows: In Section 2, the related work in area of kernel methods is discussed. Some popular applications and drawbacks of these methods are also presented. Section 3 presents the details of designing and choosing of kernel methods. In Section 4, the new proposed design for kernel method is discussed. Section 5 includes the discussion and experimental results details. Section 6 concludes the results obtained in comparison to other popular methods implemented here with the proposed work.

II. RELATED WORK AND APPLICATIONS

There are many least publicly known accomplished applications of kernel methods [10][13]. Roughly outlining three general areas of kernel method application that uses positive definite kernels are following:

•Scattered data fitting (deterministic and stochastic).
•Numerical solution of partial differential equations (deterministic and stochastic).
•Statistical or machine learning and classification, finance, multivariate integration or optimization, engineering design, computer graphics, signal processing.

A new method using Kullback Leibler (KL) measure with kernel density is used to estimate breast cancer disease probabilistically[1]. Receiver operating characteristics ROC analysis has been used to estimate performance of kernel method. KL method was tested to be 98.3% accurate in diagnosis and efficient as compared to existing methods.

Before 2001 the mapping with discrete structures to the feature space is fixed before the learning by choice and has to remain same throughout the learning [2]. A kernel function to be used must be carefully chosen otherwise non-optimally chosen kernel will give sparse representations and overbalance the benefits [2]. A new non-linear feature extraction algorithm using kernel discriminative common vectors (KDCV) kernel is used which outperform several existing face recognition methods [3].

A hybrid neural network model which uses possible combination of different transfer function and kernel function in hidden layer of feed forward neural network is presented [4]. Choice of most adequate projection function strongly depends on dataset tackled.Kernel trick is used to obtain class conditional pdf for corresponding training vectors in higher dimensional space [5].

In short this technique may be used in any supervised classifier which uses class conditional pdf. In future other kernels can be used like catchy inner product, polynomial, power exponential and inverse multi quadratic. An architecture of feed forward kernel neural network (FKNN) is proposed that can include large family of existing feed forward neural network and meets most practical requirements [6].A special kernel principal component analysis (KPCA) is used here so that all hidden layers of such network are not required be tuned and their parameters may be randomly given values and can independently be used for training data.

To detect doctored images attempts have been made. So using K-LDA produces 61.2 % accurate results [9]. Here back propagation neural network and kernel LDA is used. Here a family of fast, flexible, lightly parameterized and general purpose kernel learning derived from fast food basis function expansions [11]. Such kernels inference with the learning cost O(n) for any n training points produces predictions cost O(1) per test point. Many recent approaches have demonstrated that more expressive kernel functions which can discover rich data structures in data without human intervention can be developed [12].
Non parametric flexibility of kernels when combined with structural properties of deep neural network create scalable expressive model which forms scalable deep kernel. There is a powerful link between kernel trick and RKHs and exponential family [16].

A family of generative kernels mostly used with exponential family distribution that may outperform other kernels by the used of probability distributions [14]. Universal kernels can be constructed explicitly which are defined on compact metric spaces [10]. It also showed expressive kernel learning with improved predictive accuracy [8]. There are many more applications in real world.

Some of fast kernels are mentioned in Table 1. Below given table Table2 has list of some of popular kernel methods. Kernels are also used for multi category classification, ranking and ordinal regression. Table3 shows some of popular performance measure, computations achieved data structure and data sets used with kernel methods. Kernel methods are also utilized in statistical modeling. Parametric exponential models can be used by defining a statistical model via reproducing kernel hilbert space (RKHS) which uses the one generating kernel k. Process that transform a linear algorithm into a more general kernel method is known as kernelisation. Kernel normalization is used when diverse range of data is used so that to avoid data in larger numeric ranges dominating those in smaller ranges. There are two concepts that underlie most kernel methods: the kernel trick and the represented theorem.

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)
Vol 5, Issue 2, February 2018

TABLE 1: NEW FAST KERNELS

Kernel function	Formula	Application Area
GMK(Gaussian Mixtures Kerenl)	K(x,x')=k(x+q,x'+q)	Improverd predictive performance with efficiency
PWLK(Piece Wise Linear kernel)	Depend on piece wise function used	Used with Polynomial regretsion
KISSGP(Kernel Interpolation for scalable Structured gaussian process)	Depend upon choice of scalable structured used .	Used for near exact accurate performance in approximation.Used for natural sound modelng

TABLE 2: POPULAR KERNELS

Kernel function	Drawback	Application area	Formula
Linear kernel (Vapnik 1995)	Do not support multidimesions, Used with vector data	Functional analysis,Quantum mechaniccs	$K(x,x')=z(x)^T z(x')$
Polynomial kernel (Vapnik 1995)	May have conditional numerical instability .	Natural language processing,Basketmining ,Mapping computations,Inverted indexing .	$K(x,y)=(x^T y+c)^d$
RBF (Radial Basis Function)kernel (Vapnik 1995)	Sometime do not scale well in high dimensional data.	Support Vector Machine(SVM) classification , uses Euclidean distance as notion of similarity	$K(x,x')=z(x)z(x')$
Spline kernel (Gunn 1998)	For even value of "p" Bp is not a kernel	In computer aided design ,computer aided manufacture	$K(x,x')=B_{2p+1}(x- x')$
Sigmoid kernel (Evgenion 2000)	this is not always a positive definite	Used with two layer sigmoidal neural network	$k(x,x')=\tanh(kx^T x'+c)$
String kernel (Vishwanathan and Small 2004)	Used for string only ,can not handle high dimensionality	Natural language processing using dynamic programming	$K(x,y)=w_s \#_s(x)\#_s(x')$

A. KERNEL METHOD GOVERNING CONCEPTS

Kernel functions are mostly used for the functional, numerical and stochastic analysis based on the concepts such as [7][14][15]:

1. REPRODUCING KERNEL HILBERT SPACE[(RKHS)1950]I:-t is Hillbert space of function in which point evaluation is a continuous linear functional. A Hilbert functional space H_k for kernel k can be defined as the set of function f : X—>R for n > 0 and $X=\{x_1, x_2, \ldots, x_n\}$.The value of f(x) of a function f at a point x can be written as dot product in H_k as follows f(x)=(f , k(x,.)) and so taking f(.)=k(x',.) reproducing property valid for any x , x' is k(x , x')=(k(x,.),k(x',.)). So functional space H_k is known as reproducing kernel hilbert space (RKHS).
2. SOBOLEV SPACE:-It is a vector space of functions equipped with a norm that is a combination of L^p norm of function itself and its derivatives up to a given order. It is space of function with sufficiently many derivatives for some application domain.
3. HILBERT SCHMIDT INTEGRAL OPERATORS:-It is a type of integral transform. It is both continuous and compact.
4. GREEN"S KERNEL:-It is the impulse response of an inhomogeneous differential equation defined on a domain with specified initial conditions or boundary conditions.
5. CONVERGENCE ANALYSIS:-It is the process of some functions and sequences approaching a limit under certain conditions.
6. MERCER'S THEOREM(1909):-It is a representation of a symmetric positive definite function on a square as a sum of a convergent sequence of product function.
7. GAUSSIAN RANDOM FIELDS(GRF):-It is a random field involving Gaussian probability density functions of the variables.A one dimensional GRF is also known as Gaussian process.

B. VALID KERNELS AND ITS CHARACTERISTICS

Kernel must satisfy the Mercer's condition. Or the kernel functions earlier more precisely referred as the Mercer kernels [15]. Kernels which become right class of kernels have the following properties: A function k : X x X —>R
1.Symmetry:-such that k(x, x')=k(x',x) or the $K_{ij}=K_{ji}$.
2.Positive definite :- 1<=i , j<=n and $c^T k\, c>=0$ for any c belong to R^n.
3.Continuous :- these can have an integral form
4.Closed Convex:-the set of kernels is closed convex cone that is if c1,c2>=0 then c1.k1+c2.k2 is positive definite and if $k(x, x')=\lim_{n \to \text{infinity}} K_n(x, x')$ exists for all x ,x' then k is positive definite kernel.

In particular these methods are closed under sum, direct sum, product, and tensor product, multiplication by a scalar, zero extension, exponentials and point wise limit.

III. DESIGNING AND CHOOSING KERNEL METHOD

Any algorithm which uses positive definite kernels will surely want to choose best kernel to work with or design it. Ways to choose the best kernel are:-
1. So ideally any kernel method resulting function giving minimum expected risk.
2. When this measure cannot be easily used most commonly used approach is to approximate it using cross validation.
Here we will name few different approaches used for designing a new kernel [8].
1)Fixed rules usage (summation or multiplication of kernel etc).
2)Heuristic approaches (using measures like kernel matrices or performance value).
3)Optimization approaches (uses parameterized combinations to get optimized results).
4)Linear combination of already existing kernels.
5)Non linear combination (like exponential, power etc).
6)Data dependent combination (using specific weights for each instance).
7)Bayesian approaches (works on Bayesian hierarchical model).
8)Boosting approaches (uses combined kernel).
9)By using translation invariant kernels and kernels on semi groups.
10)One way to convert a function into valid kernel known as empirical kernel map.
11)One step method (Using sequential or simultaneous step to form a new kernel from base kernels).
12) Two steps method (an iterative approach is used to combine base kernels).

There can be a valid kernel which performs poorly and also there exist a kernel that performs ideally. So based on these below concepts one can distinguish between bad and good kernel.
1) Completeness:-If it is able to use all information necessary to represent the concept in problem domain.
2) Correctness:-It is the extent to that the underlying semantics of the task re-obeyed in the kernel trick.
3) Appropriateness:-If polynomial mistake bounds can be derived for the algorithms using that kernel trick.

IV. PROPOSED NEW DESIGN FOR KERNEL METHOD

Here we are going to give a algorithm for designing a new kernel from the existing .As per above study we need an ideal kernel to get accurate and optimized results. With above all points in mind a new algorithm is designed which will remove the problem of inefficiency and give optimized results.

New kernel design Algorithm steps are:
1. Select features to work upon .Normalize by scale up or scale down factor as per current problem need.
2. Do the novelty or outlier detection if needed after normalization. As in some problems it may not be needed.
3. The new obtained kernel formed by taking log of existing kernel should follow equivalence relation which proves it to be an ideal kernel.
4. Perform dot product and get the final results.

V. DISCUSSION AND EXPERIMENT RESULTS

To know the ease of kernel method implementation as a new user to kernel methods. Only few standard kernel methods are coded and run using Matlab software. Random data set of size 100 and 500 are taken. Than polynomial data is used for linear regression, Gaussian kernel, linear kernel and the RBF kernel to show which method gives best data matching.

TABLE 3: STANDARD DEVIATION OF VARIOUS KERNEL METHODS

Function Used	Standard Deviation Value for random data size 100	Standard Deviation Value for random data size 500
gaussian kernel	1.45	0.48
linear kernel	4.1	2.8
linear regression	1.154	0.29
RBF kernel	1.41	0.49

Table 3 has standard deviation values for data size 100 and 500 for the Gaussian kernel ,linear kernel ,linear regression and RBF kernel .Less is the value of standard deviation better is the method .Table 4 shows the

ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)
Vol 5, Issue 2, February 2018

execution time taken for data matching. Lesser the time taken faster is method as consuming less of time .

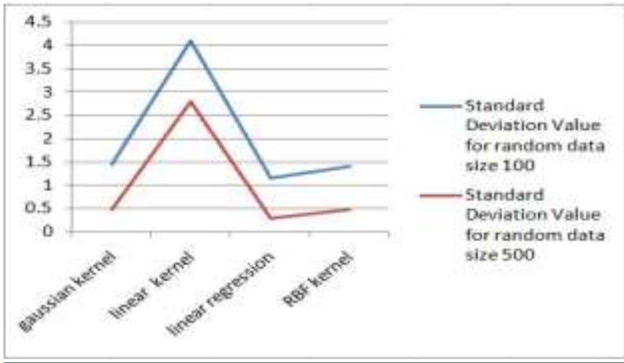

Figure 1 :shows the data matching for different methods with different colors for data size of 100.

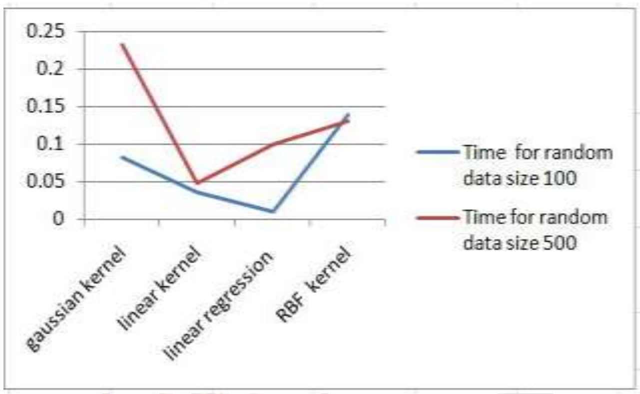

Figure 2: Time taken by Standard kernel for data size 100 and 500 the data matching for different methods with different colors for data size of 100 and 500.

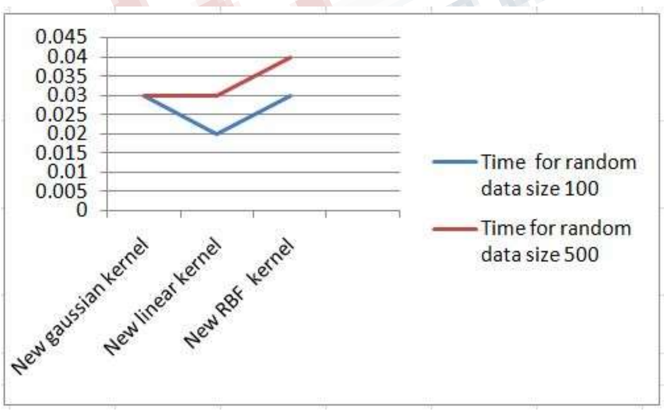

Figure 3: Time for random dataset size 100 and 500 taken by new kernels

From the above results we can say that kernel regression has less standard deviation than normal or linear regression. As the data size increases the standard deviations also decreases.And we can say accuracy of predictions is improved.

TABLE 4: EXECUTION TIME OF VARIOUS KERNELS METHODS

Function Used	Time for random data size 100	Time for random data size 500
gaussian kernel	0.082	0.232
linear kernel	0.036	0.048
inear regression	0.01	0.10
RBF kernel	0.14	0.13

In the light of above descriptions and explanations we can highlight the point that kernel trick achieved various performance criteria such as scalability, efficiency and accuracy as compared to their existing counter methods. Kernel trick are simple to implement.

Table 5 has results of new kernel methods implementation .We get less time for same data by new kernel method as compared to standard kernel method results. It also highlights the fact kernel are flexible and easy to implement so different kernel tricks can be obtained by various transformations.

A dedicated Kernel trick can be used to handle both types of variables that are continuous and categorical variables. These methods run faster than their existing other methods. Choosing the kernel particularly is the matter of geometry of the embedded data. Gaussian kernel tends to be effective on wide range of data sets as long as variance is properly set.

TABLE 5: EXECUTION TIME OF VARIOUS NEW KERNELS METHODS

Function Used	Time for random data size 100	Time for random data size 500
New gaussian kernel	0.03	0.03
New linear kernel	0.02	0.03
New RBF kernel	0.03	0.04

Choosing the kernel particularly is the matter of geometry of the embedded data. Gaussian kernel tends to be effective on wide range of data sets as long as variance is properly set. Tree kernel has better accuracy than normal kernel using dot products. Clinical kernel has been proved to outperform a liner kernel for heterogeneous data classification. Spine kernels can model a wide range of functions available.

VI. CONCLUSION

Above experimental result shows that kernel methods can be scalable as well as flexible. Existing learning algorithm can adapt the appropriate kernel parameter in accordance with their learning ways and the problem at hand . One of advantage of kernel method is that it can automatically select the best basis function for a task. As it can work according to the data.

By using kernel trick the quadratic or polynomial time computations are reduced to linear time computations.

In this paper step by step study is done to understand kernel existence than figure out steps needed for design of new kernel method. A new proposed kernel algorithm gets reduced time when executing on same dataset. So we can say good results can be obtained by various transformations on standard kernels resulting in a new kernel..

A lot more scope and vitality is there in kernel method learning .We need to perform many more experiments to confirm different aspects .Here only one direction is taken to work upon the design by logarithmic transformation.

In the end we would like to summarises about kernel methods strength and vitality by highlighting that kernel method may indeed are becoming " A big technology".

REFERENCES

1. Korkmaz, S.A. and Korkmaz, M.F., "A new method based cancer detection in mammogram textures by finding feature weights and using Kullback–Leibler measure with kernel estimation" in Optik-International Journal for Light and Electron Optics, 126(20), pp.2576-2583. 2015.

2. Menchetti, S., Costa, F., Frasconi, P. and Pontil, M., "Wide coverage natural language processing using kernel methods and neural networks for structured data" in Pattern Recognition Letters, 26(12), pp.1896-1906.2005.

3. Jing, X.Y., Yao, Y.F., Yang, J.Y. and Zhang, D., "A novel face recognition approach based on kernel discriminative common vectors (KDCV) feature extraction and RBF neural network" in Neurocomputing, 71(13), pp.3044-3048. 2008.

4. Gutiérrez, P.A., Hervás, C., Carbonero, M. and Fernández, J.C., "Combined projection and kernel basis functions for classification in evolutionary neural networks" in Neurocomputing, 72(13), pp.2731-2742.2009.

5. Sama, A., Angulo, C., Pardo, D., Català, A. and Cabestany, J., "Analyzing human gait and posture by combining feature selection and kernel methods" in Neurocomputing, 74(16), pp.2665-2674. 2011.

6. Gopi, E.S. and Palanisamy, P., "Neural network based class-conditional probability density function using kernel trick for supervised classifier" in Neurocomputing, 154, pp.225-229.2015.

7. Vert, J.P., Tsuda, K. and Schölkopf, B., "A primer on kernel methods" in Kernel Methods in Computational Biology, pp.35-70. 2004.

8. Gönen, Mehmet, and Ethem Alpaydın. "Multiple kernel learning algorithms." in Journal of Machine Learning Research: pp.2211-2268.12.Jul (2011)

9. Vinoth, S. and Gopi, E.S, "Neural Network Modeling of Color Array Filter for Digital Forgery Detection Using Kernel LDA" in Procedia Technology, 10, pp.498-504. 2013.

10. Gärtner, T., Le, Q.V. and Smola, A.J., "A short tour of kernel methods for graphs" in Under Preparation. 2006.

11. Yang, Z., Smola, A.J., Song, L. and Wilson, A.G., "A la carte-learning fast kernels" in arXiv preprint arXiv:1412.6493. 2014.

12. Wilson, A.G., Dann, C., Lucas, C. and Xing, E.P., "The human kernel" in Advances in Neural Information Processing Systems (pp. 2854-2862).2015.

13. Strobl, E.V. and Visweswaran, S., "December. Deep multiple kernel learning" in Machine Learning and Applications (ICMLA), 2013 12th International Conference on (Vol. 1, pp. 414-417). IEEE. 2013.

14. Agarwal, A. and Daumé III, H., "Generative Kernels for Exponential Families" in AISTATS (pp. 85-92). 2011.

15. Hofmann, T., Schölkopf, B. and Smola, A.J., "Kernel methods in machine learning" in The annals of statistics, pp.1171-1220.2008.