Volume 6, Issue 2, February 2016 ISSN: 2277 128X
International Journal of Advanced Research in Computer Science and Software Engineering
Research Paper
Available online at: www.ijarcsse.com

New Approach TOPAX for Optimizing Structured Data Process Based on Max-Plus Algebra

Dr. Mohamed Abdullah Mohamed *
Lecture, Dept. of Management Information, System & Higher Institute of Qualitative Studies,
Heliopolis, Cairo, Egypt

Abstract— ***This paper presents a novel approach TOPAX for a modified approach based on new concept of the classification process field. This concept is called structural risk minimization introduced by Vapnik. he proposed technique integrates between the version space technique and Topsis. Sequential hybridization enhances classification data by minimizing generalization error, determining the best structure of machine learning, in addition guarantee upper bound of classification error. Also a comparison between version space technique and the proposed technique shows the proposed technique has proved its proficiency in the two dimension of machine learning measures accuracy and speed of classification compared with version space technique.***

Keywords— Classification, Topsis, Vapnik Theory, Structural Data.

I. INTRODUCTION

Machine learning is the domain of Artificial Intelligence which is concerned with building adaptive computer systems that are able to improve their competence and/or efficiency through learning from input data or from their own problem solving experience [8]. Learning denotes changes in the system that are adaptive in the sense that they enable the system to do the same task or tasks drawn from the same population more effectively the next time [7]. The two complementary dimensions for machine learning are considered. Competence system is improving its competence if it learns to solve a broader class of problems, and to make fewer mistakes in problem solving. Efficiency system is improving its efficiency to solve the problems from its area of competence faster or by using fewer resources.

II. TIKHONOV REGULARIZATION (TRUE RISK)

Minimizing the empirical risk can lead to numerical instabilities and bad generalization performance. A possible way to avoid this problem is to restrict the class of admissible solutions, for instance to a compact set. This technique is introduced by Tikhonov and Arsenin [6] for solving inverse problems and since has been applied the learning problems with great success. In statistics, the corresponding estimators are often referred to as shrinkage estimator [5].

2.1 Vapnik-Chervonenkis Theory

Vapnik-Chervonenkis (VC) theory represents new concept considering revolution in machine learning field, this concept is structural risk minimization versus classical statistics concept empirical risk.

Definition 1 The empirical risk is defined as:

$$R_{\text{emp}}[f] := \int_{\mathcal{X}\times\mathcal{Y}} c(x, y, f(x))p_{\text{emp}}(x, y)dxdy = \frac{1}{m}\sum_{i=1}^{m} c(x_i, y_i, f(x_i)).$$

Definition 2 The structure risk is defined as:

$$R_{\text{reg}}[f] := R_{\text{emp}}[f] + \lambda\Omega[f].$$

Here λ > 0 is the so-called regularization parameter which specifies the trade-off between minimization of Remp[f] and the smoothness or simplicity which is β enforced by small $\Omega[f]$ Usually one chooses $\Omega[f]$ to be convex, since this ensures that there exists only one global minimum, provided R emp[f] is also convex. They prove the following inequality that describes an upper bound of classification error based on the structure of machine learning for unseen data.

$$\varepsilon(f) \le \varepsilon_n(f) + \sqrt{\frac{h(\log(\frac{2n}{h})+1) - \log(\frac{\eta}{4})}{n}} \quad [2]$$

Where $\varepsilon(f), \varepsilon_n(f)$ denotes the true error and empirical error. The probability that this bound is crashed equals η, machine learning capacity (VC dimension) of classifiers is h. VC confidence is the second term (the square root) of the inequality (1) [2]. It is worth mentioning that these two dimensions were well known in the seventies of the last century. The main dilemma of the mathematical dimension is the inability to find a method clear and specific on how to measure the VC - dimension of various machine learning. Inequality (1) reflected the general shape but failed to explain how to

find the VC - dimension to each machine learning. So, it could not answer many questions about the practical challenges in applying the theory. This deficit has weakened many of the importance of this theory in practice and application.
The constructive dimension which is extended to the mathematical dimension. This dimension is concerned with taking a suitable complexity model to fit the data to attain an acceptable bound of the error through determining the VC - dimension for different types of machine learning. One of the Pioneers in the constructive dimension is Girosi [3] who presents a new technique for estimating VC dimension bounds based on the theory of statistical learning. In the same direction Kon and Raphael [4] enhance the paper of Girosi [3] for finding the L1-norm of an upper bound function. Later, Kon and Raphael [5] develop the Girosi's approach [3] for existing approximate bound of the error based on Hilbert spaces. This method is called Reproducing Kernel Hilbert Spaces. Key et al. [6] who pay their attention for determining the bound of the error using Bayesian decision–theoretic trend. Also, Miller [7] presents a new mechanism for selecting the variables in the model of regression that minimizes the approximate bound of the error. In the same direction, Teytaud and Lallich [8] display the mechanism of using an appropriate VC-dimension to be a restricted bound on the accepted risk from the database. This method is named “association rules”.
This area attracts many researchers who contribute various methods for estimating the upper bound based on the theory of statistical learning. They enhance the accuracy of the bound by presenting a new concept called Rademacher's complexity but this method allows the curse of dimensionality [9]. Vapnik and Chervonenkis [10] present the new concept of Structural Risk Minimization (SRM) which displays a trade-off between the complexity of the bias function and the accuracy of the approximation. Also, Onshuus, and Usvyatsov [11] propose a method to give the uniform bound on VC. In the same trend, the uniform bound on VC technique has more attention through the work of Shelah et al. [12].

III. NEURAL NETWORK AS LEARNING AGENT

Neural networks are predictive models loosely based on the action of biological neurons. a neural network has to be configured such that the application of a set of inputs produces (either 'direct' or via a relaxation process) the desired set of outputs. Various methods to set the strengths of the connections exist. One way is to set the weights explicitly, using a priori knowledge. Another way is to 'train' the neural network by feeding it teaching patterns and letting it change its weights according to some learning rule. A neural network has to be configured such that the application of a set of inputs produces. The VC-dimension of a neural net with a binary output measures its "expressiveness". The related notion of the fat-shattering dimension provides a similar tool for the analysis of a neural net with a real-valued output. Figure 1illustrates multi layers Perceptron.

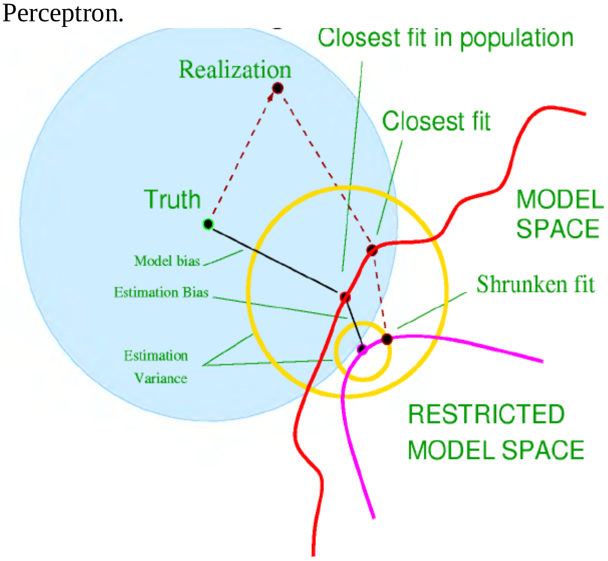


Fig1 Multi layers Perceptron.

IV. TECHNIQUES FOR OPTIMIZING GENERALIZATION ABILITY OF MACHINE LEARNING

4.1 Version Space

Version space learning algorithm provides more functions than the original learning algorithm. In the original version space, sets S and G are defined as a learning algorithm which is modified to provide these additional functions. In suggestion method, the hypotheses in S/G no longer necessarily include/exclude all the positive/negative training instances presented so far, since noisy and uncertain data may be present. Instead, the most consistent I hypotheses are maintained in the S set and the most consistent j hypotheses are maintained in the G set. Another development appears in version space that called cover set. A guarding problem can naturally be modeled as a set system (U, S) in which the universe U of elements is the set of points we need to guard and our collection S of sets contains, for each potential guard g, the set of points from U seen by g.

4.2 Regularization In Matrix Relevance Learning

A regularization method is extended recently introducing Generalized Matrix LVQ. This learning algorithm extends the concept of adaptive distance measures in LVQ to the use of relevance matrices. In general, relevance learning can display a tendency towards over-simplification in the course of training. An overly pronounced elimination of dimensions in feature space can have negative effects on the performance and may lead to instabilities in the training. Complementing the standard GMLVQ cost function by an appropriate regularization term prevents this unfavorable behavior and can help to improve the generalization ability [10].

V. TOPSIS

TOPSIS is a multiple criteria method to identify solutions from a finite set of alternatives based upon simultaneous minimization of distance from an ideal point and maximization of distance from a nadir point. TOPSIS can incorporate relative weights of criterion importance. This paper reviews several applications of TOPSIS using different weighting schemes and different distance metrics, and compares results of different sets of weights applied to a previously used set of multiple criteria data. The acronym TOPSIS stands for technique for preference by similarity to the ideal Solution.

VI. DATA SETS (BENCHMARKING PROBLEMS)

This subsection displays in table 1 six types of benchmark of classification that be used in the experiments.

Table 1 The data of benchmark of classification that be used in the experiments

Data set	Bench mark name	No. of Variables	No. of rows
First	Vovel	10	990
Second	Flag	3	871
Third	Dermatology	9	699
Fourth	Heptitates	19	155
Fifth	Cleveland heart disease	13	303
Sixth	Diabetes	8	766

These datasets selected used in the experiments are generated from UCI machine learning database.

6.1 Performance Measures

The accuracy level is used to measure the accuracy of the proposed approach. CPU time is a second measure to indicate the consumed time of the algorithm.

6.2 Simulation Procedure

The simulation procedure is used in the experiments based on dividing the whole data set, fifty percentages of them used for training and thirty percentages of them used for validating, twenty percentages of them used for testing. These experiments produced: 20 times repeated; the fold cross validation is 10.

1- Classify the problem by machine learning h .
2- Find the relative importance of variables at input n dimensional.
3- Construct A= [a_{ij} *] such* $1 \prec i \leq n, 1 \prec j \leq n$ *the matrix*
Whose entries are the confidence intervals of VC dimension to arc_{ij} *.*
4 -Construct b= [b_{ij} *] such* $1 \prec i \leq n, 1 \prec j \leq n$ *the matrix whose*
entries are empirical error achieved by running the arc_{ij}
5-Construct C=$[c_{ij}]$ *=A*$\otimes$ *b such* $1 \prec i \leq n, 1 \prec j \leq n$ *the matrix*
Whose entries are structural error achieved for each arc_{ij}
6-Consrtuct $C^1, C^2, C^3, \ldots\ldots\ldots\ldots\ldots\ldots\ldots, C^{n-1}$
7-Fnd $C^* = C^1 + C^2 + C^3 + \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots + C^{n-1}$
8-choose the minimum value in matrix C^* $[c_{ij}] = C_{ij}$
such$1 \prec i \leq n, 1 \prec j \leq n$
9-if $c_{ij} \leq b_{ij}$ *for* $1 \prec i \leq n, 1 \prec j \leq n$ *stop, else go to step 10.*
10 -if the minimum value $C_{ij} \leq \varepsilon$ *such* ε *is the accepted risk. Stop,*
Else n=n-1 go to step 2
Output the final hypothesis.

Fig2 Collective Structural Machine Algorithm

VII. EXPERIMENTS AND RESULTS

In this section a comparison between version space and the proposed technique is made as a group of the famous benchmark of classification in the first stage, then, the result is tested by t-test to explore if there is a significance difference between the version space technique and the proposed technique at different levels of error such as at 5%, 1% respectively. Then the comparison with two independent samples is made with different variance. So for determining degree of the freedom, the special formula is made. This section is divided into five subsections for each benchmark. The following benchmark is implemented using the following cost function.

$$L(\alpha) \le L_{emp}(\alpha) + \sqrt{\frac{h(\log\frac{2N}{h}+1) - \log\frac{\eta}{4}}{N}} \quad \text{: Error Bounds}$$

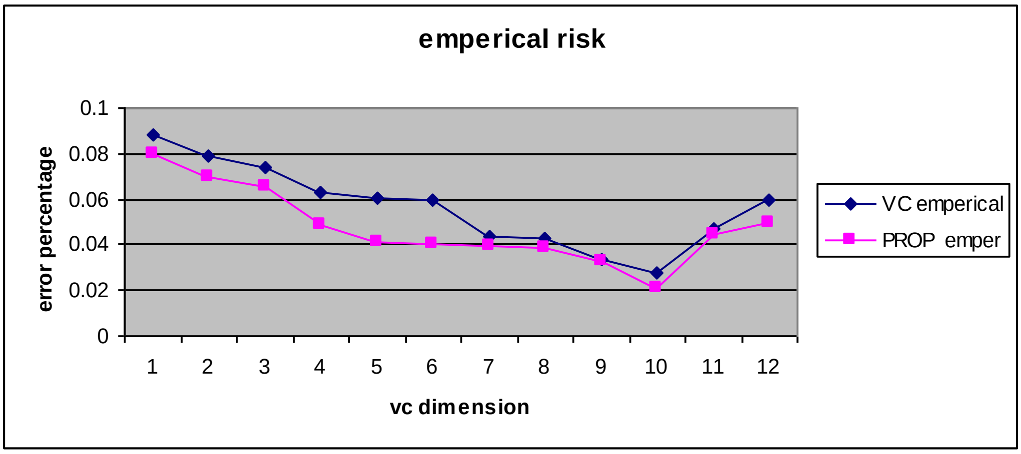
Such L(α) is the true error, L(α) emp is the empirical error, N number of data used, h is the VC capacity, η is percentage predetermined of error. From the previous formula which based on two factors N, h. It is very clear that the negative relation of the function of true error and h the VC dimension. If the number of data constant, the only factor minimize the function is h. The hypothesis with minimize h has the minimum true error (Structural Risk Minimization).

7. 1 Vovel Bench Mark

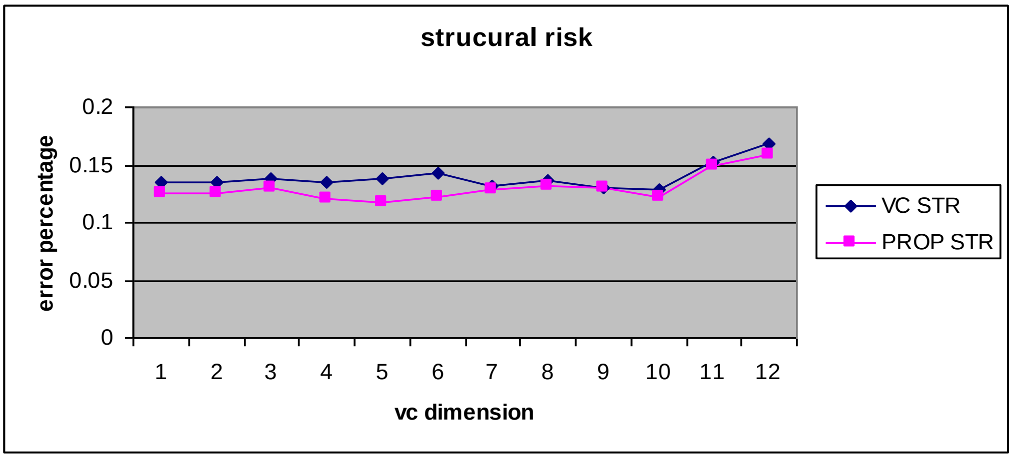
Target variable: Diagnosis. Number of predictor variables 13 Type of model: Multilayer Perceptron Neural Network (MLP) Number of layers: 3 (1 hidden). Hidden layer 1 neurons: Search from 2 to 20 Hidden layer activation function: Linear Output layer activation function: Linear Type of analysis: Classification. Category weights (priors): Data file distribution Misclassification costs: Equal (unitary) Validation method: Cross validation Number of cross-validation folds: 10% - Input Data Number of variables (data columns): 14 Data sub setting: Use all data rows Number of data rows: 9090 Total weight for all rows: 303 Rows with missing target or weight values: 0 Rows with missing predictor values: 6. **Table 1** shows the classification error (mean of 20 trails) of VOVELbench mark obtained by version space and hybrid (between version space and topsis respectively). **Graph 1** Shows the empirical classification error (mean of 20 trails) of VOVELbench mark Obtained by version space and hybrid (between version space and topsis respectively). **Graph 2** shows the structural classification error (mean of 20 trails) of VOVELbench mark Obtained by version space and hybrid (between version space and topsis respectively).

Table 2 The classification error (mean of 20 trails) of vovelbench mark

VC Dimension	No of Data	VC Error	Empirical Risk of Version Space	Empirical Risk of Proposed Technique	VC STR	PROP STR
1	**9090**	***0.045842***	**0.08861**	**0.079433**	***0.134451***	***0.125274***
2	**9090**	***0.056017***	**0.078977**	**0.069765**	***0.134994***	***0.125783***
3	**9090**	***0.064163***	**0.074255**	**0.065654**	***0.138418***	***0.129817***
4	**9090**	***0.071123***	**0.0632**	**0.049123**	***0.134323***	***0.120246***
5	**9090**	***0.07728***	**0.060447**	**0.040767**	***0.137726***	***0.118047***
6	**9090**	***0.082848***	**0.059568**	**0.040007**	***0.142416***	***0.122855***
7	**9090**	***0.08796***	**0.043899**	**0.039878**	***0.131859***	***0.127838***
8	**9090**	***0.092706***	**0.04311**	**0.038765**	***0.135816***	***0.131471***
9	**9090**	***0.09715***	**0.033257**	**0.032568**	***0.130407***	***0.129717***
10	**9090**	***0.101338***	**0.028006**	**0.018988**	***0.129344***	***0.120326***
11	**9090**	***0.105308***	**0.047437**	**0.042332**	***0.152745***	***0.14764***
12	**9090**	***0.109088***	**0.059888**	**0.049257**	***0.168976***	***0.158344***



Graph 1 The Empirical Classification Error of 20 Trails Obtained by Version Space and Hybrid



Graph 2 The Structural Classification Error of 20 Trails Obtained by Version Space and Hybrid of Version Space and Topsis

Table 3 Shows the required time of vovelbench mark

	VERSION SPACE	HYBRID VERSION
CPU-TIME	525.87S	4.19S

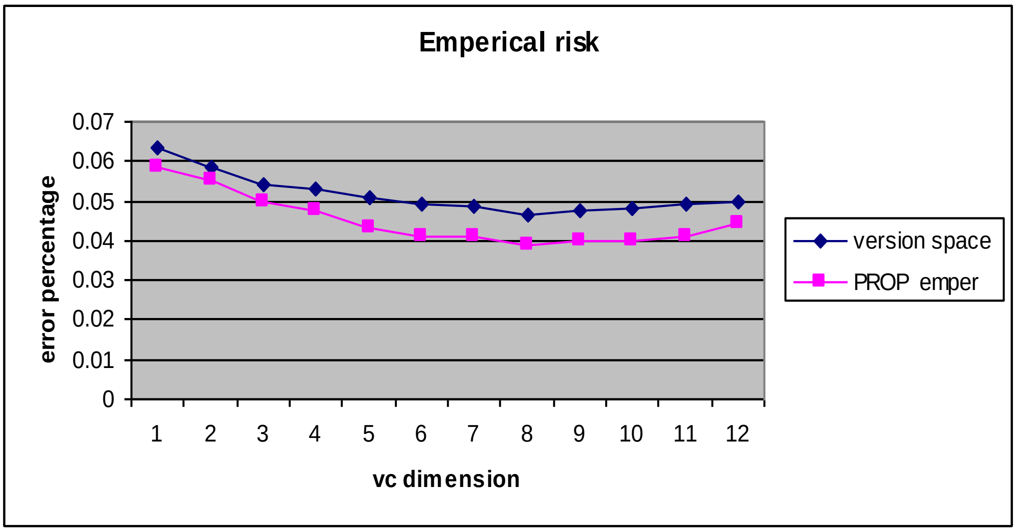
The previous results obtained we notes that minimum empirical error at the complexity of VC equal 10 to give empirical error equal to0.028006 to the version space technique, 0.018988 to proposed technique, but the true error (new concept) 0.129344, 0.120326 respectively. the classical statistics will choose the complexity of VC equal 10 to give minimum empirical error but on the other hand in the new approach we choose the complexity of VC equal 5 to give the minimum true error. The multi –Perceptron has structure complexity 5 is the optimal structure to achieving minimum true error (generalization error) *0.118047* although has empirical error 0.040767. because our target general error not empirical error. learning not memorization. on the other hand, the proposed technique gives higher speed more than version space.

7.2 Flag Benchmark

Target variable: Habitat Number of predictor variables: 20 Type of model: Multilayer Perceptron Neural Network (MLP) Number of layers: 3 (1 hidden) Hidden layer 1 neurons: Search from 2 to 20 Hidden layer activation function: Logistic Output layer activation function: Logistic Type of analysis: Classification Category weights (priors): Data file distribution Misclassification costs: Equal (unitary) Validation method: Cross validation Number of cross-validation folds: 10 Input Data. Number of variables (data columns): 23 Data subsetting: Use all data rows Number of data rows: 84170 Total weight for all rows: 383 Rows with missing target or weight values: 0 Rows with missing predictor values: Table 3. Shows the classification error (mean of 20 trails) of flag bench mark Obtained by version space and hybrid (between version space and topsis respectively.

Table 4 The classification error of 20 trails of flag bench mark obtaining by version space and hybrid (Between Version Space and Topsis)

VC dimension	no of data	VC error	empirical risk of Version space	empirical risk of Proposed technique	VC STR	PROP STR
1	84170	*0.015918*	0.063322	0.058765	*0.07924*	*0.074684*
2	84170	*0.019793*	0.058655	0.055098	*0.078448*	*0.074891*
3	84170	*0.02289*	0.054368	0.049677	*0.077257*	*0.072566*
4	84170	*0.025535*	0.052778	0.047354	*0.078314*	*0.072889*
5	84170	*0.027878*	0.050988	0.043268	*0.078866*	*0.071146*
6	84170	*0.029999*	0.049124	0.041235	*0.079122*	*0.071233*
7	84170	*0.031948*	0.048768	0.041001	*0.080715*	*0.072948*
8	84170	*0.033759*	0.046554	0.038765	*0.080314*	*0.072525*
9	84170	*0.035458*	0.047655	0.039877	*0.083113*	*0.075334*
10	84170	*0.03706*	0.047855	0.04	*0.084915*	*0.07706*
11	84170	*0.038581*	0.048956	0.041099	*0.087537*	*0.07968*
12	84170	*0.040031*	0.04999	0.044077	*0.090021*	*0.084107*



Graph3. Shows the structural classification error (mean of 20 trails) of flag bench mark

Table 5 Shows the required time of flag benchmark

	version space	**hybrid version**
CPU time	2m:33.88s	4.89s

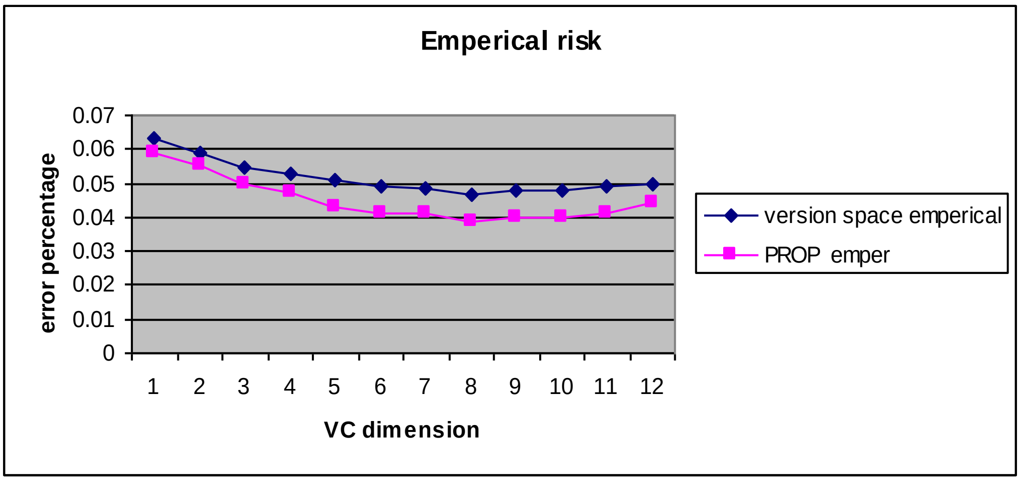
The previous results obtained we notes that minimum empirical error at the complexity of VC equal 8 to give empirical error equal to 0.046554 to the version space technique .0.038765to proposed technique, but the true error (new concept) *0.080314*, *0.072525* respectively. the classical statistics will choose the complexity of VC equal 8 to give minimum empirical error but on the other hand in the new approach we choose the complexity of VC equal 5 to give the minimum true error. The multi –Perceptron has structure complexity 1 is the optimal structure to achieving minimum true error (generalization error) *0.071146* although has empirical error 0.043268. because our target general error not empirical error. learning not memorization. On the other hand, the proposed technique gives higher speed more than version space.

7.3 Dermatology Benchmark

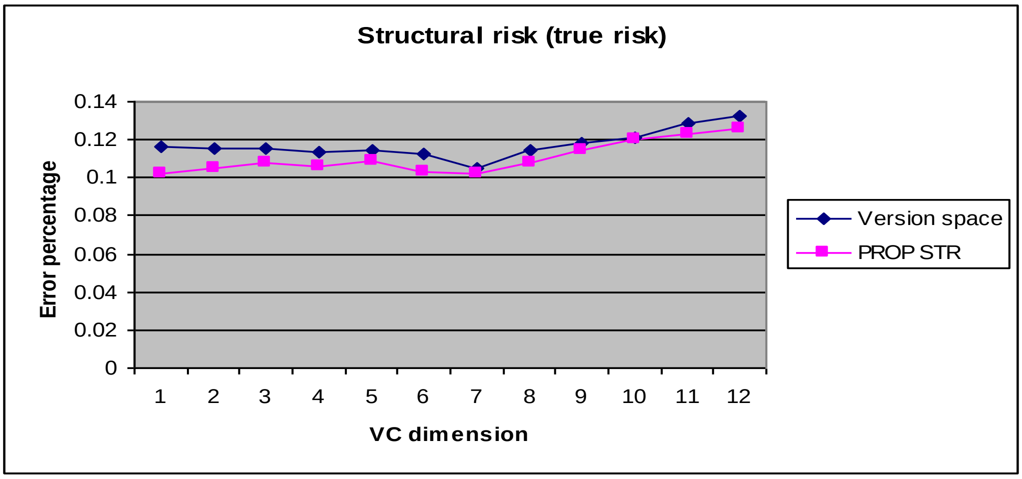
Target variable: Good/Bad Number of predictor variables: 34 Type of model: Multilayer Perceptron Neural Network (MLP) Number of layers: 3 (1 hidden) Hidden layer 1 neurons: Search from 2 to 20 Hidden layer activation function: Logistic Output layer activation function: Logistic Type of analysis: Classification Category weights (priors): Data file distribution Misclassification costs: Equal (unitary) Validation method: Cross validation Number of cross-validation folds: 10 Input Data Number of variables (data columns): 35 Data subsetting:Use all data rows Number of data rows: 351 Total weight for all rows: 10530 Rows with missing target or weight values: 0 Rows with missing predictor values: 0

Table 5 Shows the classification error (mean of 20 trails) of dermatology bench mark

VC dimension	**no of data**	**VC error**	**empirical risk of Version space**	**empirical risk of Proposed technique**	**VC STR**	**PROP STR**	**VC dimension**
0.5	10530	14.11819	***0.036616***	0.079435	0.065437	*0.116051*	*0.102053*
1	10530	19.24918	***0.042755***	0.072966	0.062457	*0.115721*	*0.105212*
1.5	10530	24.11855	***0.047859***	0.067902	0.060077	*0.115761*	*0.107935*
2	10530	28.81802	***0.052314***	0.061545	0.053688	*0.113859*	*0.106002*
2.5	10530	33.39115	***0.056312***	0.058004	0.052457	*0.114316*	***0.108769***
3	10530	37.86361	***0.059965***	0.052146	0.043579	*0.11211*	*0.103544*
3.5	10530	42.25234	***0.063345***	**0.042001**	**0.038767**	***0.105346***	***0.102111***
4	10530	46.5694	***0.066502***	0.047987	0.041344	*0.114489*	*0.107846*
4.5	10530	50.82379	***0.069474***	0.048765	0.044678	*0.118239*	*0.114151*
5	10530	55.02251	***0.072286***	0.049025	0.047666	*0.121311*	*0.119952*
5.5	10530	59.17115	***0.074962***	0.053677	0.047777	*0.128639*	*0.122738*
6	10530	63.27428	***0.077517***	0.054655	0.048348	*0.132173*	*0.125865*



Graph4. Shows the empirical classification error (mean of 20 trails) of Dermatology bench mark



Graph5: Shows the structural classification error (mean of 20 trails) of Dermatology bench mark

Table 6. Shows the required time of dermatology benchmark

	version space	hybrid version
Cpu-time	456.2s	19.81s

The previous results obtained we notes that minimum empirical error at the complexity of VC equal 3.5 to give empirical error equal to 0.042001 to the version space technique 0.038767 to proposed technique, but the true error (new concept) *0.105346*, *0.102111*respectively. the classical statistics will choose the complexity of VC equal 3.5 to give minimum empirical error but on the other hand in the new approach we choose the complexity of VC equal 2.5 to give the minimum true error. The multi –Perceptron has structure complexity 2.5 is the optimal structure to achieving minimum true error (generalization error) *0.108769* although has empirical error 0. 052457.because our target general error not empirical error. learning not memorization. on the other hand, the proposed technique gives higher speed more than version space.

VIII. CONCLUSION

From the previous results a comparison between version space and the proposed technique on group of the famous benchmark

- Cleveland Heart: the proposed technique accuracy is higher than version space. on the other dimension speed of the proposed technique is faster than version space.
- Flag benchmark: the proposed technique accuracy is higher than version space. on the other dimension speed of the proposed technique is faster than version
- Dermatology benchmark: the proposed technique accuracy is higher than version space. on the other dimension speed of the proposed technique is faster than version

The comparison which made in most cases showed the proficiency of the proposed technique sequential hybridization between version space and topsis. The results obtained give high indicator that propsed technique has prove its efficiency: maximum generaliztion ability of machine learning in other words minimum true error (generalization error) which means the propsed techinque guarantee maximum percentage of error for unseen data. Besides, it gives the optimal structure for machine learning used.

REFERENCES

[1] James King. (2008)." VC-Dimension of Visibility on Terrains "CCCG, Montréal, Quebec.
[2] Olson.D.L. (2004) "Comparison of Weights in TOPSIS Models" Mathematical and Computer Modeling. www.elsevier.com/locate/mcm.
[3] Peter L. Bartlett. (2000)."Vapnik-Chervonenkis Dimension of Neural Nets"
[4] Petra Schneider1 and Kerstin Bunte1 and Han Stiekema1 (2008)"Regularization in Matrix Relevance Learning" University of Groningen, Institute for Mathematics and Computing Science
[5] Swief R. A., Y. G. Hegazy, T. S. Abdel-Salam, and M.A Bader (2008)"Load -Price Forecasting Model Employing Machine Learning Techniques".
[6] Tikhonov. A. N. and V. Y. Arsenin. (1977). "Solution of Ill-Posed Problems". Winston, Washington, DC.
[7] Vapnik N Vladimir. (1999)" An Overview of Statistical Learning Theory" IEEE Transactions on neural networks, vol. 10, no. 5, September.
[8] Vapnik V. and Chervonenkis A. (1964.)," A note on one class of perceptrons. Automation and Remote Control,"
[9] Vapnik V. And Chervonenkis A. "Theory of Pattern Recognition." in Russian. Nauka, Moscow, (1974). (German Translation: W. Wapnik & A. Tscherwonenkis, Theorie der Zeichenerkennung, Akademie-Verlag, Berlin.
[10] Weston. J, A. Elisseeff, and B. Scholkopf. (2001)." Use of the £0-norm with linear models and kernel methods. Technical report". Biowulf Technologies, New York.
[11] Onshuus A, Usvyatsov A. On dp-minimality, strong dependence and weight, the journal of Symbolic Logic (76), 2011. 737-740.
[12] Shelah. S. Strongly dependent theories. Israel J. Math.2008. Accessed 15 July 2014. Available. http://arxiv.org/abs/math/0504197.
[13] Riondato M, Vandin F. Finding the True Frequent Itemsets. Jan 08 2013. Accessed 28 June 2014. cs. LG cs. DB cs.DS stat. ML arXiv: 1301. 1218v3.available https://scirate.com/search.
[14] Zaher H, Abdullah M, Ragaa N. A Social Learning Approach for Minimizing True Risk of Collective Machine Learning, International Journal of Advanced Research in Computer Science and Software Engineering (3), 2013; 1172-1179.
[15] Dempster A P. Upper and lower probabilities induced by a multivalued mapping. The Annals of Mathematical Statistics (38), 1967; 325–339.Doi:10.1214/aoms/1177698950.
[16] Shafer, Glenn A. Mathematical Theory of Evidence. Princeton University Press, ISBN 0-608-02508-9, 1976. Doi: 10.1090/s0002-9904-1977-14338-3.

