



A Survey of Techniques Used for Optimizing Generalization Ability of Machine Learning based on True Risk Minimization

Hegazy Zaher*

*Professor of Mathematical Statistics,
Institute of Statistical Studies and
Research (ISSR), Cairo University,
Egypt.*

Naglaa Raga Said

*Assistant Professor of Operations
Research, (ISSR).Cairo University,
Egypt.*

Mohamed Abdullah

*Dept. of Management Operations
Research, Higher Institute of qualitati
studies. Egypt.*

Abstract: *Vapnik and Chervonenkies introduce the True Risk Minimization approach in 1974. This approach is considered as a great improvement in machine learning field; this approach has its roots in a broad spectrum of various fields as machine learning, Statistics, and optimization. This paper presents a novel survey for Optimizing Generalization Ability of Machine Learning machine learning based on True Risk Minimization introduced by Vapnik and Chervonenkies (VC) . The paper displays one hundred papers that cover the art of Optimizing Generalization Ability of machine learning from 1989 to November 2013. The main criterion is used in this paper for classifying papers is the type of field considered tackling Optimizing Generalization Ability of Machine Learning based on True Risk Minimization. Fields considered this problem are Statistics, Mathematics, Computer science, Information Science, in addition, the paper represents different applications for different science.*

Key words: *Machine Learning, Statistical Learning Theory, Vapnik and Chervonenkies, True Risk Minimization.*

1- INTRODUCTION

Vapnik and Chervonenkies pay their attention for selecting a suitable complexity of model to can fit data for attaining approximately an upper bound of minimum true error. They also present a formula that gives an upper bound of the true error for unseen data They also give the following mathematical formula that describes an upper bound of the true error for unseen data (Montana et al, 2005).

$$\varepsilon(f) \leq \varepsilon_n(f) + \sqrt{\frac{h(\log(\frac{2n}{h}) + 1) - \log(\frac{\eta}{4})}{n}} \dots\dots\dots (1) \text{Where } \varepsilon(f), \varepsilon_n(f) \text{ are true}$$

error, and empirical error respectively, η is the probability that the given bound is violated, and h is a measure of machine learning capacity which is called the VC dimension of classifiers. The second term of the inequality (1) is named the VC confidence.

This Paper provides just an outline of the methodology; theoretical proofs and more detailed explanations be found in the original work (Vapnik, 1999). Up till now, there is no a general method for measuring VC dimension. This motivates many researchers to pay their attention towards how determining VC dimension for various machine learning. This paper introduce a novel survey of Techniques Used for Optimizing Generalization Ability of Machine Learning based on True Risk Minimization from 1989 to November 2013. The novelty of this work is the criterion used for the classification based on the field considered tackling this problem .the sub criterion is year of published paper into the field.

The remainder of this paper is structured as follow; Section two states published papers in the field of statistics . Section three provides published papers in the field of mathematics. Section four presents published papers in the field of Computer Science. Section five introduces published papers in the field of Information Science. Section six displays published papers in the field of Applied Sciences. Section seven presents the discussion and conclusion .

2- STATISTICS FIELD

1989 - Eric. B .Baum and David Haussler focused on the answer of the question when the researchers expect the performance of network be generalized from random training examples selected from certain probability distribution.

2002- Theodoros Evgeniou et al study the framework of Regularization Theory which mainly based on ill-posed problems.

2005- Stephan Boucheron1 et al provide a survey that represent an overview of studying the theoretical methods, and algorithms be used in generalization of machine learning.

2006- Bin Li and Prem K. Goel study the issue of the regularized model and its sensitivity based on the prior determination through the Bayesian view . They suggest a new technique based on t-tailed priors for the achieving robust Bayesian solutions.

2006- Ming Yuan and Yi Lin study and introduce efficient algorithms which are considered as extensions of these methods used for factor selection. They proved these methods give superior performance compared to the traditional methods.

2010- Sylvain Arlot presents a survey for cross-validation methods that have widespread strategy. This survey shows results obtained by the most recent work of model selection theory.

2010- Xin GAO and Peter X.-K. SONG suggest a novel method based on complicated likelihood model of the Bayes information criterion (BIC) and emphasize the property of consistency for the model selection.

2012 - Yongdai Kim et al pay their attention for asymptotic properties for how to select model within regression problems that have high-dimensional in case the sample size are less than dimension of covariates.

3- MATHEMATICS FIELD

1999- Theodoros Evgeniou and Massimiliano Pontil introduce novel technique for computation the VC dimension for regression problems that have bounded subspaces of generating Kernel Hilbert Spaces (RKHS).

2003- G'Abor Lugosi Et Al show the existence of certain class of subsets for VC dimension that has the symmetric convex hull. They answered the question was motivated by many academic literature for the Theoretical properties of algorithms be used in machine learning.

2004- Gabor Lugosi and Marten Wegkamp suggest a model selection based on penalized empirical loss minimization that used in nonparametric classification problems.

2006 -Giorgio Corani and Marino Gatto introduce a method of estimation VC-dimensions for ecological models which are nonlinear through the approach proposed by Vapnik et al. (1994).

2008- Giorgio Gnecco present a study of relationship between model complexity (which means what required number of computational units). They introduce an approximation error is examined based on tools derived from Statistical Learning Theory.

2008- Kristin P. Bennett presents the interplay of and machine learning optimization. They suggest framing for Stackelberg games that fundamental problem of machine learning problems

2008- M. Murugappan et al introduce a system of emotion recognition from EEG (Electroencephalogram) signals. The primary aim of this work is introducing a comparison for the efficacy of human emotions . This comparison is based on two discrete systems of wavelet transform (DWT) by using the feature extraction process compared to three statistical features.

2010-Tomasz Luczak And Stephan Thomass introduce new bound for the chromatic number expressed by terms of fractional transversality and combined VC-dimension approach created by Vapnik.

4- COMPUTER SCIENCE FIELD

1993- Yaser S. Abu-Mostafa used the VC dimension for establishing tools for learning analyze from generated examples. The proposed method is used for analyzing learning from hints.

1994- Vapnik presents a new method for evaluating the learning machine capacity. The novel approach depend on fitting a theoretical proof induced the function which has the measure of empirical maximal difference between the error rates for two sets of separate data that different size.

1995 -Corrina Cortes and Vladimir Vapnik introduce method for the support-vector network which is considered a new learning machine used for two-group of classification problems. The machine actually performs the novel idea: converting input vectors which are non-linearly to another feature space has very high dimension.

1996- Anthony M. Zador and Barak A. Pearlmutter Provide the following contributions firstly, it present a novel tackling for the computational capacity of this sort of dynamic system; and (2) the proposed method offer a framework used for analyzing the computational efficiency for the dynamical systems .

1996- David E. Moriarty and Risto Miikkulaine introduce a novel learning method based on reinforcement learning called SANE (Symbiotic, Adaptive Neuro-Evolution). Main idea of this method is generating a population of neurons by genetic algorithms to establish a neural network has generalization performance.

1996-Richard S. Sutton achieves the positive results for optimization tasks are controlled. He clears the fundamental differences between the proposed method and classical methods of machine learning optimization .the proposed method is called sparse-coarse-coded function approximators (CMACs).

1997- Martin Anthony presents a survey for the 'probably approximately correct' (PAC) technique of learning process and some of its variants. This survey focus on the sample complexity.

1998-Adam Krzy'zak and Tam'as Linder introduce a novel application of complexity regularization which gives an estimation bounds for the function of nonlinear depend on neural network espfific radial basis function network.

1998- Christopher J.C. Burges shows the mechanism of how Support Vector machines can have infinite VC dimension through estimating the VC dimension for homogeneous polynomial and Gaussian radial basis function kernels.

1999-Vladimir N. Vapnik introduce new approach of machine learning is called the Statistical learning theory was began in the late 1960's. Before the 1990's the theory was an absolutely theoretical background of the problem the new method of learning algorithms is called support vector machines.

2000- Walter Zucchini proved that high-dimensional sets of data that have interrelated dependency structures leads to intractable computational complexity.

- 2001- Markus Junker and Andreas Dengel** present a novel learning techniques that patterns are used for text categorization . this new method is not based on an attribute-value representation of documents .
- 2002- JACEK LEŃSKI** introduce a novel method of learning tolerant of imprecision is presented and be depend on Neuro-fuzzy modeling.
- 2002- Olivier Bousquet and Andre Elisseeff** introduce a definition notions used for stability of learning algorithms .in addition the mechanism of using these notions to explore error bounds of generalization depend on the leave-one-out error and the empirical error.
- 2003- Juha Reunanen** studies a fundamental methodological flaw which is used in comparison of methods for variable selection . in addition, he presents a empirical guide for the computation performance of cross-validation evaluates various variable subsets.
- 2003-Yoshua Bengio and Nicolas Chapados** present Metric-based methods that have recently been used for the regularization process. This method often generating very significant development based on cross-validation.
- 2004- Trevor Hastie Et al** prove the support vector machine (SVM) is a common tool used for classification .they introduce different efficient methods for modeling two-class .
- 2004- Vorontsov K. V.** presents a Combinatorial cross-validation method that determine for learning algorithms the generalization performance. In addition ,he introduce an Upper bounds are explored that are better than those generated in the theory of Vapnik– Chervonenkis .
- 2005- Vikas Sindhwani vikass** et al introduce The enormous wealth for the data unlabeled in various applications of machine learning. These methods is corner stone for posing challenges to model methods of semi-supervised learning.
- 2005-Will Bridewell et al** present a novel review for the paradigm of inductive process modeling, that uses as background knowledge for constructing quantitative models used for dynamical systems.
- 2005- Zakria Hussain and John Shawe-Taylor** present an analysis for matching pursuit based kernel principal components analysis (KPCA) .
- 2006- Guo-Zheng Li et al** create a novel algorithm which is called GA-MTL (genetic algorithm based multi- task learning) is suggested to specify the features used for the input and/or the target of NNE.
- 2006-Lars Graning, Yaochu Jin, and Bernhard Sendhoff** study the generalization improvement in field of multi-objective learning (MO). In addition, a case study is introduced for neural network classifiers the generation of depend on the receiver operating characteristics (ROC) .
- 2006- Kristin P. Bennett and Emilio Parrad Hern´andez focus on** the fields of machine learning , in addition mathematical programming. They tackle the Optimization problems .
- 2006- Lorenzo Rosasco et al** present a huge class of regularization techniques modeled for creating solutions of ill-posed inverse problems generates rise to new learning algorithms.
- 2006- Mikhail Belkin et al** suggest a class of learning algorithms depend on a novel structure of regularization .this class allows their to exploit marginal distribution geometry.
- 2006- Pak-Ming Cheung pakming and James T. Kwok** study kernel methods used for formulti-instance learning. They introduce Experiments on both regression and classification problems prove that the suggested method gives improved performance
- 2006- Uehua Cui et al.** suggest an interval-mapping technique for phenotypes measured in counts. They present a model for the effects of QTL by modeling a generalized Poisson regression .44
- 2007-Deok-Hwan Kim, et al** introduce a relevance feedback technique depend on multi-class for the support vector machine and other sort of cluster-merging .
- 2007- Gavin C. Cawley and Nicola L. C. Talbot** study optimization model parameters for kernel machine in convex optimization problem.
- 2007- Gideon S. Mann and Andrew McCallum** present an expectation regularization for the semi-supervised learning technique based on the family of exponential parametric .
- 2007- Leonardo Vanneschi** studies the generalization problems of GP based on Pareto multi-optimization through the training process that gives better generalization ability than classical GP .
- 2007- Miroslav Dudík et al** study the generalization problems of the a fractional function .they describe different applications in machine learning field as robust Fisher linear discriminant analysis, and signal processing .
- 2007-Yinyin Liu, et al** discover novel approach is called an optimized approximation algorithm (OAA). This approach is suggested to tackle the overfitting problem in function approximation based on neural networks (NNs).
- 2007- Yann Guermeur** focused on studying Vapnik’s statistical learning theory. He introduces classifier that has large margin multi-category .
- 2008- Gregory Druck et al** propose a technique for training models that have discriminative probabilistic with unlabeled instances and labeled features .
- 2008- Inhoi Koo and Rhee Man Kil** introduce a novel technique of selection a model for regression machine learning problems through the modulus of continuity.
- 2008- Kuzman Ganchev** introduces a posterior regularization of machine learning type that have a probabilistic framework and weakly supervised learning.
- 2008- Partha Pratim Talukdar et al** present a novel approach is called Alternating True Optimization (ASO) . this approach can enhance the supervised learning process through learning feature representations from data are unlabeled.

- 2008- Robert Burbidge** study the generalization ability of genetic programming that is to automatically explore computer programs used in solving problems.
- 2008- van der Maaten et al** present a comparative study for the techniques used in dimension reduction of nonlinear problem .
- 2008 - Vikas Sindhwani** presents a novel approach based on co-training between machine learning for semi-supervised kernel methods.
- 2008- Yaochu Jin, Bernhard Sendhoff** study the machine learning field through the concept of a multi objective task. The proposed technique is based on the Pareto- multiobjective optimization approach.
- 2009 - Chuong B. Do and Quoc V. Le** Introduce novel online and batch algorithms used for the training process a various supervised learning models (such as SVMs, logistic regression 60
- 2009- Frank-Michael Schleifl and Thomas Villmann** study the area of machine learning that related to neural maps and Learning Vector Quantizer .they has success to achieve the simplicity and robustness in many application .
- 2009- Marcin Owczarczuk** introduce a novel class for support vector machines used for the binary classification task. This class has superior performance for support vector machines which is established by using only two support vectors .
- 2009 Yanyan LAN*et al** improve a theoretical framework for the process of ranking, and prove the mechanism of how to make a generalization analysis of list wise ranking algorithms based on the proposed the framework.
- 2010- Gavin C. Cawley and Nicola L. C. Talbot** create strategies for Model selection of machine learning algorithms basically combine the numerical optimization of a suitable model selection .
- 2010- Martin Jaggi and Marek Sulovsky** suggest a novel approximation algorithm for building the recent sparse approximation SDP solver of (Hazan, 2008).
- 2010- Michael W. Mahoney et al** improve the procedure for extracting useful information generated from noisy data through adding some kinds of norm constraint to optimize the modified objective function.
- 2010-Stephen Boyd et al** study the area of convex optimization for many problems in the field of statistics and machine learning.
- 2010-Tong Zhang** study the learning formulations related to the non-convex objective functions which is used in practical applications. This work a try to tackle the gap between theory field and practical field . 68
- 2010-Tristan Fletcher et al** study the Multiple Kernel Learning (MKL)which is used for replicating the signal combination process. This process control rules when they collect multiple sources of information .
- 2011- Lin Xiao** focuses on the area of regularized stochastic learning, in addition online optimization problems. Where this objective function of model is the sum of two convex terms.
- 2011- Zeeshan Ahmed and Saman Majeed** introduce methods for tackling the process of optimization for best results estimation using the approach of adaptive machine learning.
- 2012- Carl Brunner et al** study the problem of generalization Pairwise classification. They present new method of generalization based on kernel function for using symmetric training of data sets through the framework of support vector machines.
- 2012- Chia-Hua Ho and Chih-Jen Lin** focus on the study of classification of support vector machines and the regression of Support vector and how generalization process is often time consuming.
- 2012-Dean Foster** studies the generalization problem in case of a test distribution is different from the training distribution. He introduce a technique for selecting more features than domains at the same time avoiding Overfitting problem through utilizing properties of variance for data-dependent.
- 2012 -Dugald Macpherson, and Sergei Starchenko** focus on generalization problem in case of training data set structure are smaller number of sample domains.
- 2012- Emilie Morvant Emilie and Liva Ralaivola** suggest new method called a PAC-Bayes bound using for the generalization error of the Gibbs classifier through framework of multi-class classification. The creation in work is the using *confusion matrix* for the classifier as a measure of error .
- 2012- James Bergstra and Yoshua Bengio** represent the Grid search and manual search which be the most common used tools for optimization the generalization process .
- 2012 - Kiri L. Wagstaff** develops the current machine learning (ML). his research proved that machine learning lost its communication as problems imported to the real world of science
- 2012- Kiri L. Wag staff** introduces six Impact Challenges to explicitly concentration for the field's energy . in addition , he presents the attention, and discussion for existing obstacles that should be studied.
- 2012-Le Song et al** present a new framework for the selection feature depend on maximization dependence between the chosen features and the labels of an evaluation problem based on the Hilbert-Schmidt Independence Criterion.
- 2012- Matthias Aschenbrenner et al** focus on the Vapnik-Chervonenkis (VC) density of predetermined families for specific theories of stable first order.
- 2012- Neil D. Lawrence** present a novel technique for spectral dimensionality reduction based on the Gaussian Markov random fields (GRFs). The proposed method is called maximum entropy unfolding (MEU) using for a nonlinear generalization performance of principal component analysis.
- 2012- Olcay Taner Yildiz** states and proves the lower bounds of the VC-dimension for the case of univariate decision tree hypothesis category.
- 2012- Ran El-Yaniv and YairWiener** prove existence a strong relation between two common learning techniques: the first called stream-based active learning and the second perfect selective classification
- 2012 -Sham M. Kakade et al** analyze the systematic method for establishing matrix-based regularization techniques .

2012-Trinh-Minh-Tri Do and Thierry Artieres study the Machine learning as an optimization problem based on convex objective function to rely on optimizers of convex

2013-Hegzy Zaher, Naglaa Said, Mohamed Abdullah present a novel approach of optimizing generalization ability based on Social machine learning. The novel approach used tropical algebra for achieve the social learning .

2013- Matteo Riondato and Fabio Vandin prove that Frequent item of sets mining is a basically primitive in data mining.

2013- Tanmoy Chakraborty focuses on the generalization performance of the writing behaviors of individuals and unique linguistic styles.

5 -INFORMATION SYSTEM SCHOOL

2004- Ying Tan, and Jun Wang introduce a novel mechanism for training support vector machines (SVMs) that have a hybrid kernel and low Vapnik- Chervonenkis (VC) dimension.

2009- Farshad Kyoomarsi et al study the problem of proliferation the Internet and the huge amount of transferred data.

2009-Xiaowen Zhao et al tackle the generalization performance for the landslide prediction based on GIS through constructing the landslide prediction model depend on SVM (support vector machine).

2011-Christopher Lee study generalization performance of information theory would present useful metrics for statistical inference.

2012- Eugen Pircalabelu and Gerda Claeskens study the two specific classes of graphical models, the first called Directed Acyclic Graphs (DAG) and Gaussian Graphical Models (GGM).

6- APPLIED SCIENCE

6-1 Mathematical Psychology

2000- Walter Zucchini introduces a novel model selection that addressed for non-professional who have primary knowledge of the statistical concepts .

2006- Sébastien Hélie display how to develop the generalization ability of Model selection is used psychology field. he develops the principals of generalization machine learning **to be suitable in the applications of** psychology.

6-2 Psychonomic

2008- Frank Jäkel et al present a novel method is called exemplar models that actually generalize very well.this method is based on Kernel methods .

6-3 Finance

2009- Akinori Hirabayashi et al suggest a generalization the Genetic Algorithm (GA) system to can generate trading rules depend on Technical Indexes.

6-4 Economics

2010 - Henry de-Graft Acquah introduce a comparison of performance for the two commonly used techniques in model selection process. Firstly, Akaike information criteria (AIC) and secondly ,Bayesian information criteria (BIC) for discriminating between price transmission techniques restricted various conditions.

6-5 Biomedical Engineering.

2012- Chen Tao and Hong Zeng-lin represented a new selective SVM based on KFCM that is suggest to enhance the generalization ability.

6-6 Management Science

2012-ZHAO Dan presents that accumulated a huge number of reviews of customer for goods and website services can be the Support vector machine (SVM) an effecting text categorization method.

6-7 Medicine

2012- P. K. Srimani* and Manjula Sanjay Koti develop the generalization of designing high-performance for the computer-aided diagnosis systems, enhancing the accuracies of the machine learning algorithms .

7-DISCUSSION & CONCLUSION

This paper shows that one hundred papers of literature review that concerned with optimizing generalization ability of machine learning based on True Risk Minimization offered by a various sciences, statistics, mathematics, Computer Science, information science, and also introduces varieties of application in different fields. The survey shows that 71% of the papers published in Computer Science School. 8% published in Statistics school, 8% published in mathematics school.5% published in information system.8% published in applications of optimizing generalization ability of machine learning in various application medicine, economics, finance, and psychology.

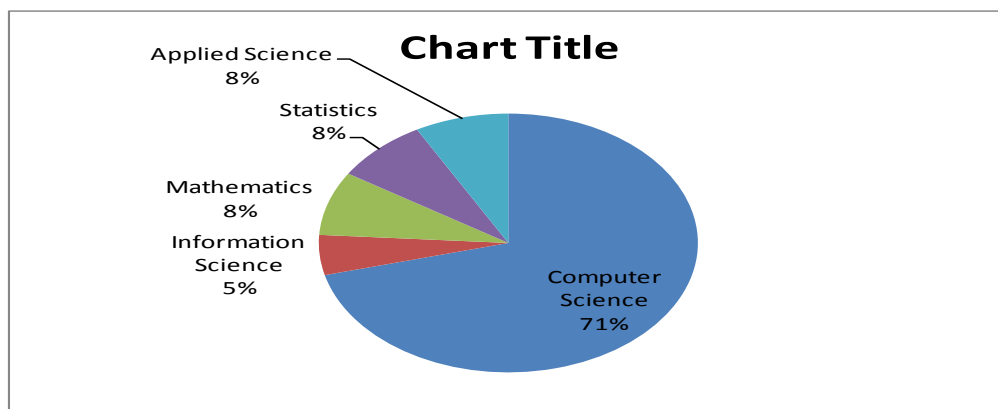


Fig1: illustrates the percentage of contribution each field for Optimizing Generalization Ability of Machine Learning based on True Risk Minimization

REFERENCES

- [1] Adam Krzyżak, and Tamás Linder,(1998). *Radial Basis Function Networks and Complexity Regularization in Function Learning*. IEEE Transactions On Neural Networks, Vol. 9, No. 2, March 1998. 247.
- [2] Akinori Hirabayashi et al (2009) *Optimization of the Trading Rule in Foreign Exchange using Genetic Algorithm*. GECCO'09, July 8–12, 2009, Montréal, Québec, Canada.
- [3] Anthony M. Zador and Barak A. Pearlmutter .(1996).*VC Dimension of an Integrate-and-Fire Neuron Model*.To appear (1996) in *Neural Computation* 8(3).
- [4] Bin Li and Prem K. Goel. (2006). Regularized Optimization In Statistical Learning: A Bayesian Perspective. *Statistica Sinica* 16(2006), 411-424.
- [5] Carl Brunner et al.(2012).Pairwise Support Vector Machines and their Application to Large Scale Problems . *Journal of Machine Learning Research* 13 (2012) 2279-2292 Submitted 8/11; Revised 3/12; Published 8/12
- [6] Chen Tao and Hong Zeng-lin (2012) "*A Selective SVM Ensemble Algorithm based on KFCM*".2012 International Conference on Mechanical Engineering and Automation Advances in Biomedical Engineering, Vol.10 .
- [7] Chia-Hua Ho and Chih-Jen Lin.(2012)."*Large-scale Linear Support Vector Regression*". *Journal of Machine Learning Research* 13 (2012) 3323-3348 Submitted 4/12; Revised 10/12; Published 11/12.
- [8] Christopher J.C. Burges.(1998)."*A Tutorial on Support Vector Machines for Pattern Recognition*". *Data Mining and Knowledge Discovery*, 2, 121–167 (1998). 1998 .
- [9] Christopher Lee (2011) "*Empirical Information Metrics for Prediction Power and Experiment Planning*".*Information* 2011, 2, 17-40; doi:10.3390/info2010017
- [10] Chuong B. Do and Quoc V. Le.(2009)."*Proximal regularization for online and batch learning*".Appearing in *Proceedings of the 26 th International Conference on Machine Learning*, Montreal, Canada, 2009.
- [11] Corrina Cortes and Vladimir Vapnik. (1995)."*Support-Vector Networks*".*Machine Learning*, 20, 273-297 (1995)
- [12] Davide Anguita et al.(2000) ."*Evaluating the Generalization Ability of Support Vector Machines through the bootstrap*". *Neural Processing Letters* 11:51-58,2000
- [13] David E. Moriarty and Risto Miikkulaine . (1996)."*Efficient Reinforcement Learning through Symbiotic Evolution*". *Machine Learning*, 22, 11{33 (1996) c 1996 .
- [14] Deok-Hwan Kim, et al (2007) ."*Support Vector Machine Learning for Region-Based Image Retrieval with Relevance Feedback*". *ETRI Journal*, Volume 29, Number 5, October 2007.
- [15] Emilie Morvant Emilie and Liva Ralaivola .(2012)."*PAC-Bayesian Generalization Bound on Confusion Matrix for Multi-Class Classification*", published in "*International Conference on Machine Learning 2012, Edinburgh : United Kingdom (2012)*"
- [16] Eric B. Baum, David Haussler (1988)."*What size nets give valid generalization*". Technical report of Computer Research Laboratory, University of California, Santa Cruz. [0] ESAIM: PS ESAIM.(2005)
- [17] Farshad Kyoomarsi.(2009)."*Optimizing Machine Learning Approach Based on Fuzzy Logic in Text Summarization*".*International Journal of Hybrid Information Technology*.Vol.2, No.2, April, 2009
- [18] Frank Jäkel et al .(2008)."*Generalization and similarity in exemplar models of categorization: Insights from machine learning*". *Psychonomic Bulletin & Review* 2008, 15 (2), 256-271.applied science . *Psychonomic*
- [19] Frank-Michael Schleiße and Thomas Villmann.(2009) "*Neural Maps and Learning Vector Quantization -Theory and Applications*". proceedings, *European Symposium on Artificial Neural Networks - Advances in Computational Intelligence and Learning*. Bruges (Belgium), 22-24 April 2009,
- [20] GABOR LUGOSI et al (2003) . *A Note On The Richness Of Convex Hulls Of Vc Classes* . Submitted 2 May 2003 , accepted in final form 17 November 2003.AMS 2000 Subject classification: 62G08, 68Q32.
- [21] Gavin C. Cawley and Nicola L. C. Talbot "*Preventing Over-Fitting during Model Selection via Bayesian Regularization of the Hyper-Parameters*". *Journal of Machine Learning Research* 8 (2007) 841-861 Submitted 11/06; Published 4/07

- [22] Gavin C. Cawley and Nicola L. C. Talbot (2010) "On Over-fitting in Model Selection and Subsequent Selection Bias in Performance Evaluation". *Journal of Machine Learning Research* 11 (2010) 2079-2107 Submitted 10/09; Revised 3/10; Published 7/10 .
- [23] Gideon S. Mann and Andrew McCallum "Simple, Robust, Scalable Semi-supervised Learning via Expectation Regularization". Appearing in *Proceedings of the 24 th International Conference on Machine Learning, Corvallis, OR, 2007.*
- [24] Giorgio Corani and Marino Gatto.(2006)."VC-dimension and True Risk Minimization for the analysis of nonlinear ecological models".*Applied Mathematics and Computation.*
- [25] Giorgio Gnecco.(2008)" Approximation Error Bounds via Rademacher's Complexity". *Applied Mathematical Sciences, Vol. 2, 2008, no. 4, 153 – 176 .*
- [26] Gregory Drucker al .(2008) . " Learning from Labeled Features using Generalized Expectation Criteria". SIGIR'08, July 20–24, 2008, Singapore
- [27] Henry de-Graft Acquah(2010) "Comparison of Akaike information criterion (AIC) and Bayesian information criterion (BIC) in selection of an asymmetric price relationship". *Journal of Development and Agricultural Economics Vol. 2(1) pp. 001-006, January, 2010.*
- [28] Imhoi Koo and Rhee Man Kil.(2008) ." Model Selection for Regression with Continuous Kernel Functions Using the Modulus of Continuity". *Journal of Machine Learning Research* 9 (2008) 2607-2633 Submitted 1/07; Revised 8/08; Published 11/08 .
- [29] JACEK ŁE,SKI.(2002)." Improving The Generalization Ability Of Neuro-Fuzzy Systems By "-Insensitive Learning". *Int. J. Appl. Math. Compu T. Sci., 2002, Vol.12, No.3, 437–447.*
- [30] James Bergstra and Yoshua Bengio.(2012)." Random Search for Hyper-Parameter Optimization" .*Journal of Machine Learning Research* 13 (2012) 281-305 Submitted 3/11; Revised 9/11; Published 2/12.
- [31] Joseph B. KADANE and Nicole A. LAZAR." Methods and Criteria for Model Selection ". © 2004 American Statistical Association.*Journal of the American Statistical Association .March 2004, Vol. 99, No. 465, Review Article*
- [32] Juha Reunanen .(2003) ." Overfitting in Making Comparisons Between Variable Selection Methods". *Journal of Machine Learning Research* 3 (2003) 1371-1382 Submitted 5/02; Published 3/03.
- [33] Kiri L. Wagstaff (2012) "Machine Learning that Matters".*Jet Propulsion Laboratory, California Institute of Technology, 4800 Oak Grove Drive, Pasadena, CA 91109 USA* Appearing in *Proceedings of the 29 th International Conference on Machine Learning, Edinburgh, Scotland, UK, 2012.*
- [34] Kristin P. Bennett1 (2008)." Bilevel Optimization and Machine Learning."J.M. Zurada et al. (Eds.): WCCI 2008 Plenary/Invited Lectures, LNCS 5050, pp. 25–47, 2008. Springer-Verlag Berlin Heidelberg 2008.
- [35] Kristin P. Bennett and Emilio Parrado-Hernández. (2006)"The Interplay of Optimization and Machine Learning Research". *Journal of Machine Learning Research* 7 (2006) 1265–1281 Submitted 7/06; Published 7/06
- [36] Kuzman Ganchev et al .(2009)." Posterior Regularization for Structured Latent Variable Models". *Journal of Machine Learning Research* 10 (2009) .
- [37] Lars Graning, Yaochu Jin, and Bernhard Sendhoff.(2006) ."Generalization Improvement in Multi-Objective Learning."2006 International Joint Conference on Neural Networks Sheraton Vancouver Wall Centre Hotel, Vancouver, BC, Canada July 16-21, 2006.
- [38] Leonardo Vanneschi et al . "Multi-Optimization Improves Genetic Programming Generalization Ability". *GECCO'07, July 7–11, 2007, London, England, United Kingdom. .ACM 978-1-59593-697-4/07/0007.*
- [39] Le Song(2012)"Feature Selection via Dependence Maximization" *Journal of Machine Learning Research* 13 (2012) 1393-1434 Submitted 5/07; Revised 6/11; Published 5/12 .
- [40] Lin Xiao .(2011)." Dual Averaging Methods for Regularized Stochastic Learning and Online Optimization computer science". *Journal of Machine Learning Research* 11 (2010) 2543-2596 Submitted 5/09; Revised 3/10; Published 10/10.
- [42] Lorenzo Rosasco et al .(2006)." Spectral Methods for Regularization in Learning Theory"*Journal of Machine Learning Research* () Submitted 12/05; Published 2006 computer science
- [43] Marko Robnik-Šikonja(2004) ." Improving Random Forests ". *Machine Learning , ECML 2004 Proceedings, Springer, Berlin, 2004*
- [44] Markus Junker and Andreas Dengel. (2001)."Preventing Overtting in Learning Text Patterns for Document Categorization ": *ICAPR 2001, LNCS 2013, pp. 137–146, 2001. computer science*
- [45] Martin Anthony.(1997) ." Probabilistic Analysis of Learning in Artificial Neural Networks :The PAC Model and its Variants". *Neural Computing Surveys* 1, 1-47, 1997, <http://www.icsi.berkeley.edu/~jagota/NCS>.
- [46] Martin Jaggi and Marek Sulovsky .(2010)"A Simple Algorithm for Nuclear Norm Regularized Problems" Appearing in *Proceedings of the 27 th International Conference on Machine Learning, Haifa, Israel, 2010. Copyright*
- [47] Matteo Riondato and Fabio Vandin.(2013)" Controlling False Positives in Frequent Itemsets Mining through the VC-Dimension". *arXiv:1301.1218v1 [cs.LG] 7 Jan 2013*
- [48] Michael W. Mahoney et al (2011). "Implementing Regularization Implicitly Via Approximate Eigenvector Computation". Appearing in *Proceedings of the 28 th International Conference on Machine Learning, Bellevue, WA, USA, 2011.*

- [49] Mikhail Belkin et al. (2006). "Manifold Regularization: A Geometric Framework for Learning from Labeled and Unlabeled Examples". *Journal of Machine Learning Research* 7 (2006) 2399-2434 Submitted 4/05; Revised 5/06; Published 11/06
- [50] Ming Yuan and Yi Lin. (2006). "Model selection and estimation in regression with grouped variables". *J. R. Statist. Soc. B* (2006) 68, Part 1, pp. 49–67. *statistician*.
- [51] Miroslav Dudík et al. (2007). "Maximum Entropy Density Estimation with Generalized Regularization and an Application to Species Distribution Modeling". *Journal of Machine Learning Research* 8 (2007) 1217-1260 Submitted 4/06; Revised 4/07; Published 6/07.
- [52] Murugappan. M. et al (2008) ." Time-Frequency Analysis of EEG Signals for Human Emotion Detection" N.A. Abu Osman, F. Ibrahim, W.A.B. Wan Abas, H.S. Abd Rahman, H.N. Ting (Eds.): *Biomed 2008, Proceedings* 21, pp. 262–265, 2008
www.springerlink.com © Springer-Verlag Berlin Heidelberg 2008
- [53] Neil D. Lawrence (2012) "A Unifying Probabilistic Perspective for Spectral Dimensionality Reduction: Insights and New Models ." *Journal of Machine Learning Research* 13 (2012) 1609-1638 Submitted 10/10; Revised 1/12; Published 5/12
- [54] Olcay Taner Yilidiz. (2012). "On the VC-Dimension of the Univariate Trees "ICPRAM 2012 - International Conference on Pattern Recognition Applications and Methods.
- [55] Olivier Bousquet. Andr e Elisseeff ."Stability and Generalization". *Journal of Machine Learning Research* 2 (2002) 499-526.
- [56] Pak-Ming Cheung pakming and James T. Kwok. (2006). "A Regularization Framework for Multiple-Instance Learning" Appearing in *Proceedings of the 23 rd International Conference on Machine Learning*, Pittsburgh, PA, 2006.
- [57] Partha Pratim Talukdar et al. (2008). "DRASO: Declaratively Regularized Alternating True Optimization". Appearing in the *Workshop on Prior Knowledge at the 25 th International Conference on Machine Learning*, Helsinki, Finland, 2008.
- [58] Richard S. Sutton. (1996). "Generalization in Reinforcement Learning: Successful Examples Using Sparse Coarse Coding". *Advances in Neural Information Processing Systems* 8, pp. 1038-1044, MIT Press, 1996.]
- [59] Robert Burbidge. (2008). "Genetic Programming and Support Vector Machines" .A Contribution to the *Foundations of AI: Robert Burbidge Workshop/Summer School on Evolutionary Computing* .Lecture Series by Pioneers, August 18-22, 2008, Londonderry computer science
- [60] S bastien H lie. (2006) ." An Introduction to Model Selection: Tools and Algorithms". *Tutorials in Quantitative Methods for Psychology* 2006, Vol.2 (1), p. 1-10. *applied science* . *Psychology*
- [61] Sham M. Kakade et al "Regularization Techniques for Learning with Matrices". *Journal of Machine Learning Research* 13 (2012) 1865-1890 Submitted 12/10; Revised 10/11; Published 6/12 computer science
- [62] Srimani P. K. and Manjula Sanjay Koti. (2013). "Medical Diagnosis Using Ensemble Classifiers - A Novel Machine-Learning Approach". *Journal of Advanced Computing*. (2013) 1: 9-27 doi:10.7726/jac.2013.1002
- [63] Stephan Boucheron1, et al. (2005). "Theory Of Classification: A Survey Of Some Recent Advances". *Probability and Statistics* November 2005, Vol. 9, p. 323–375 DOI. Received June 18, 2004. Accepted September 12, 2005.
- [64] Stephen Boyd et al. (2010) "Distributed Optimization and Statistical Learning via the Alternating Direction Method of Multipliers". *Foundations and Trends in Machine Learning* .Vol. 3, No. 1 (2010) 1–122
- [65] Sylvain Arlot† (2010) "A survey of cross-validation procedures for model selection" *Statistics Surveys* Vol. 4 (2010) 40–79 . [66] Tanmoy Chakraborty. (2013). "Authorship Identification in Bengali Literature: a Comparative Analysis". arXiv:1208.6268v4 [cs.CL] 24 Feb 2013.
- [67] Theodoros Evgeniou, and Massimiliano Pontil. (1999). "On the VC Dimension for Regression in Reproducing Kernel Hilbert Spaces". *A.I. Memo No. 1656* May 1999. *C.B.C.L Paper No. 172*.
- [68] Theodoros Evgeniou et al. (2002). "Regularization and statistical learning theory for data analysis". *Computational Statistics & Data Analysis* 38 (2002) 421–432 .www.elsevier.com/locate/csa
- [69] Tomasz Luczak And St_Ephan Thomass. (2010). "COLORING DENSE GRAPHS VIA VC-DIMENSION" "Mathematics Subject Classification. Primary: 05C35, Secondary: 05C15, 05C60, 68Q32.
- [70] Tong Zhang. (2010). "Analysis of Multi-stage Convex Relaxation for Sparse Regularization". *Journal of Machine Learning Research* 11 (2010) 1081-1107 Submitted 5/09; Revised 1/10; Published 3/10.
- [71] Trevor Hastie Et al (2004). "The Entire Regularization Path for the Support Vector Machine". *Journal of Machine Learning Research* 5 (2004) 1391–1415
- [72] Trinh-Minh-Tri Do and Thierry Artieres. (2012). "Regularized Bundle Methods for Convex and Non-Convex Risks". *Journal of Machine Learning Research* 13 (2012) 3539-3583 Submitted 4/10; Revised 4/12; Published 12/12 .
- [73] Tristan Fletcher et al (2010) "Currency Forecasting using Multiple Kernel Learning with Financially Motivated Features." This work is closely related to a presentation titled Multiple Kernel Learning on the Limit Order Book given at the *Workshop on Applications of Pattern Analysis* 2010.
- [74] Vikas Sindhwani vikass et al." Linear Manifold Regularization for Large Scale Semi-supervised Learning". Appearing in *Proc. of the 22 st ICML Workshop on Learning with Partially Classified Training Data*, Bonn, Germany, August 2005.

- [75] Uehua Cui, et al (2006) . "On the Generalized Poisson Regression Mixture Model for Mapping quantitative Trait Loci With Count Data ". *Genetics* 174: 2159–2172 (December 2006) .copyright _ 2006 by the Genetics Society of America.
- [76] van der Maaten et al.(2008)." Dimensionality Reduction :A Comparative Review". MICC, Maastricht University, P.O. Box 616, 6200 MD Maastricht, The Netherlands. January 2008. Preprint submitted to Elsevier.
- [77] Vladimir N. Vapnik.(1999) ."An Overview of Statistical Learning Theory.'" *IEEE Transactions On Neural Networks*, Vol. 10, No. 5, September 1999.
- [78] Vladimir Vapnik et al. (1994)" Measuring the VC –dimension of learning machine." *Neural Computation*.
- [79] Vorontsov K. V." Combinatorial Substantiation of Learning Algorithms." Dorodnitsyn Computing Center, Russian Academy of Sciences, ul. Vavilova 40, Moscow, 119991 Russia e-mail: voron@ccas.ru Received January 30, 2004. computer science
- [80] Walter Zucchini.(2000)." An Introduction to Model Selection". *Journal of Mathematical Psychology* 44, 41_61 (2000)
- [81] Will Bridewell willb, and Narges Bani Asadi nargesb.(2005)." Reducing Overfitting in Process Model Induction". Appears in *Proceedings of the Twenty-Second International Conference on Machine Learning*, Bonn, Germany, 2005.
- [82] Xiaowen Zhao et al . (2009)." Research On Landslide Prediction Model Based On Support Vector Model". *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, Vol. 38, Part II.
- [83] Xin GAO and Peter X.-K. SONG(2010)." Composite Likelihood Bayesian Information Criteria or Model Selection in High-Dimensional Data". *Journal of the American Statistical Association* December 2010, Vol. 105, No. 492, Theory and Methods.statiticians
- [84] Yaochu Jin, Bernhard Sendhoff (2008) "Pareto-Based Multi objective Machine Learning: An Overview and Case Studies". *IEEE Transactions On Systems, Man, And Cybernetics—Part C: Applications And Reviews*, Vol. 38, No. 3, May 2008
- [85] Yann Guermeur .(2007)."VC Theory of Large Margin Multi-Category Classifiers". *Journal of Machine Learning Research* 8 (2007) 2551-2594 Submitted 1/07; Revised 8/07; Published 11/07 computer science
- [86] Yaser S. Abu-Mostafa. (1993) "Hints and the VC Dimension." .*Neural Computation* 5,278-288 @ 1993 Massachusetts Institute of Technology.
- [87] Ying Tan, and Jun Wang (2004). A Support Vector Machine with a Hybrid Kernel and Minimal Vapnik-Chervonenkis Dimension". *IEEE Transactions On Knowledge And Data Engineering*, Vol. 16, No. 4, April 2004
- [88] Zhao Dan (2012) "Method for Customer Review Category Based MKL-SVM". *Management Science and Engineering* Vol. 6, No. 1, 2012, pp. 64-68
- [89] Zakria Hussain and John Shawe-Taylor(2005)
- [90] "Structure learning using a Focused Information Criterion in Graphical Models." KU Leuven - ORSTAT and Leuven Statistics Research Center, Naamsestraat 69, 3000 Leuven, Belgium.statiticians.
- [91] Dean Foster, et al (2012)." Domain Adaptation: A Small Sample Statistical Approach" Appearing in *Proceedings of the 15th International Conference on Artificial Intelligence and Statistics (AISTATS) 2012*, La Palma, Canary Islands. Volume 22 of *JMLR*:
- [92] Ran El-Yaniv , Yair Wiener (2012)"Active Learning via Perfect Selective Classification". *Journal of Machine Learning Research* 13 (2012) 255-27