Feature Selection based on Particle Swarm Optimization and Support Vector Machine

[1]Vikesh Yadav, [2]Jyoti Vashishtha
[1]Research Scholar, [2]Associate Professor
[1,2]Department of Computer Science and Engineering,
[1,2]Guru Jambheshwar University of Science and Technology, Hisar, India

Abstract: Feature selection is very important part of the classification process. In this, only that data is selected which is actually needed. Recently, with the intensification of data dimensionality, a number of algorithms of feature selection confronts in terms of efficiency and effectiveness. Feature selection is done by using a classifier along with a optimization technique. There are many optimization techniques used in feature selection such as particle swarm optimization, genetic algorithms, ant colony optimization etc. For the classification purpose, support vector machine, decision trees techniques can be used. The basic goal of this paper is to give an overview of support vector machine and particle swarm optimization techniques used for feature selection. This overview will help to discover new support vector machine and particle swarm optimization techniques based strategies which can be used to find best feature subset selection.

IndexTerms **- Feature Selection, Support Vector Machine, Particle Swarm Optimization.**

I. INTRODUCTION

In these days, a big quantity of data is accumulated in the data repositories. A big breach exists between stored data and the knowledge to be fabricated from this data. There is no automatic occurrence of this transition [1]. Here data mining comes into light. Data mining is a process to analyse data from many perspectives and then summarize it in valuable knowledge. In data analysis, on an exploratory base, some internal knowledge may be known about data, but data mining may help in extracting more in deep knowledge about data. It allows businesses to be proactive. Data mining mines and provides knowledge about future trends and customers' behaviour. In turn, business decisions are knowledge-driven, rise in revenue and improvement in costs. However, exponential intensification in dimensionality of data has put challenges to majority of present mining and learning algorithms. Here, feature selection has proved as an effective and efficient way to prepare high dimensional data for machine learning and data mining techniques [2]. Feature selection filters those features that are redundant and not related to the task [3]. It is the process of selecting features automatically which are more useful for the problem we are working on [4]. It, not only improves the quality of model, but also makes modelling process more efficient. During building model, superfluous columns consume more memory and CPU time. Even if resources are not an issue, we still perform feature selection and identify the best columns, as quality of model may be degraded by unneeded columns.

Feature selection process has four key steps which are feature stopping criterion, subset generation, subset evaluation and result validation. The feature subset generation is heuristic search process that results in selection of a candidate which is subset for evaluation. It uses searching strategies like sequential, complete, or random search for generating subsets of features. Dunne et al. proved that basis of these searching strategies is stepwise deletion or addition of features [5]. An evaluation criterion is used for goodness of generated subset. Only, in case, newly generated subset is better than previous subset, the previous subset may be replaced with new best subset. These two processes are repeated again and again until stopping criterion is met. The final best feature subset is validated by former knowledge or by using many tests. The process of feature selection is illustrated in Figure 1.

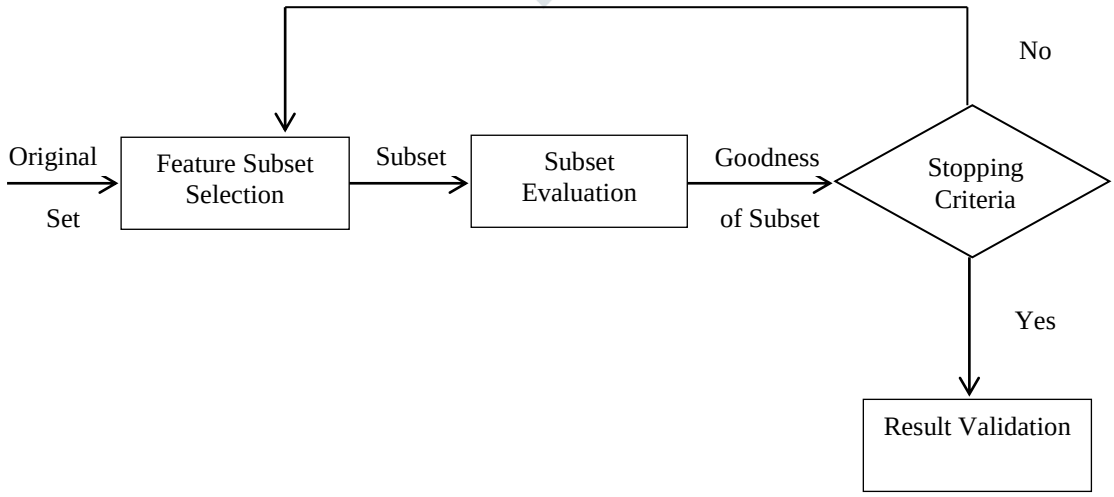

Figure 1. Feature Selection Process

II. SUPPORT VECTOR MACHINE

Support vector machine (SVM) is one of the latest learning methods used for classification. Initially a training algorithm was developed that maximise the margin between training pattern and decision boundaries. The complexity of the problem was matched by adjusting the parameters in the algorithm. The solution to a problem was expressed as a linear combination of supporting patterns. The performance of this algorithm was tested on hand written digit reorganization problems [6].

In SVM linear classifiers, a classification model is constructed by using the training data and then this model is used for making predictions [7]. It can be used for both classification and regression challenges though it is mostly used in classification problems. In this algorithm, some data may be plotted like a point in n-dimensional space with value of some feature. It may be for being value of a typical coordinate. After this we can perform distribution by finding hyper-plane. Figure 2. Illustrates the support vectors divided by a hyper plane.

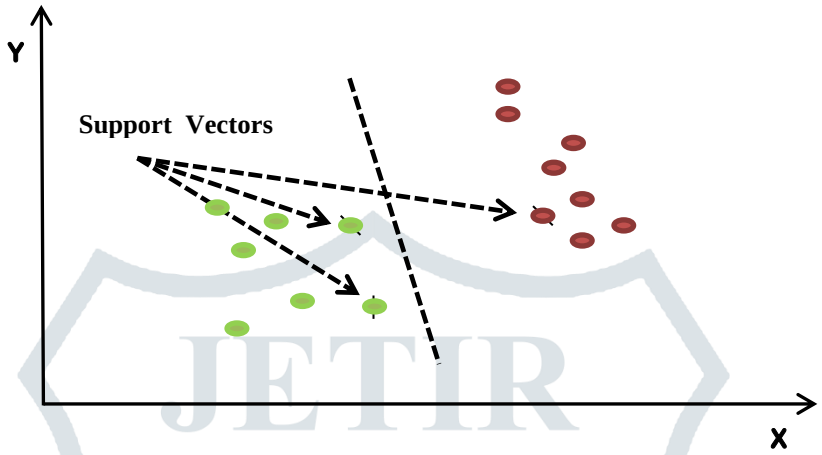

Figure 2. Support Vectors Divided By A Hyper Plane

The SVM has the capability to capture large data sets, as it is a method for structural risk minimization (SRM) that identifies the solutions with lowest risk association [8]. Hua and Sun have used SVM to predict the subcellular localization of proteins from their amino acid compositions [9]. In order to mine the protein sequences from nucleotide sequences, this is an important step to identify the points at which regions start that code for proteins. These points are called translation initiation sites (TIS). Zien et al established the applications of SVMs for identifying these TIS [10]. Cai et al introduced SVM to predict the protein structural class [11]. Burbidge and Buxton introduced SVM as a robust tool for many data mining aspects like classification, regression and outlier detection. The SVM is based on firm statistical and mathematical foundations concerning generalisation and optimisation theory. The successive improvements in kernel methods and maximum margin methods make SVM as an essential tool for data miners [12].

SVMs can be used for solving many real world problems with increased productivity and coding efficiency. Their application may reduce need of labelled training instances significantly. Image classification can be performed by using SVMs. Results experimentally depicts that SVMs accuracy is significantly higher than old query refinement schemes. SVM can identify hand-written characters. A novel classification approach for online handwriting recognition is presented by Jin Hyung Kim et al [13]. Chen et al introduced a new approach for extracting rules by using the information that is provided by dividable hyper plane and support vectors to advance the simplification capability and unambiguousness of the classification rules and decrease computational difficulty. This overcomes the problems of how to clarify the computed solutions and present the extracted information by using the SVM [14]. Barakat and Bradley assess the quality of rules, which are extracted directly from the support vectors (SVs) of a trained SVM using a modified sequential covering algorithm, using area under the receiver operating characteristics (ROC) curves (AUC) [15]. Barakat and Bradley presented another utilization of SVM which changes the black box form of an SVM into a comprehensible illustration of SVM analytic result. The outcome on diabetes data set showed that the understandable SVMs provide a capable tool for predicting diabetes. The developed rules are medically accurate and agree with the outcome of the medical studies [16]. Sun et al. presented L1-L2 SVM, which provide regression analysis with automatic feature selection for prognostic prediction. L1-L2 enhanced for prognostic prediction by using the information of censored data as constraints [17]. SVM is a frontier which divides input data into two classes. SVM is used by Basari et al for binary classification in opinion mining in movie reviews that partitions the data sets into two classes. These two classes of reviews are positive and negative classes. The positive reviews go to positive class and negative reviews go to negative class. SVM is used with 10 fold cross validation and confusion matrix. PSO is used for parameter optimization problem for selecting the best input selection feature. The results show increased accuracy of classification [18].

This algorithm has applied broadly in biological and other sciences. They have differentiated proteins up to 90% of compounds and correctly classified. Many permutation tests based on SVM weights have advised as a mechanism for the explanation of SVM models. Support vector machine weights are used for analyzing SVM models in past. By setting the SVM with good parameter and using the features rightly, the accuracy of the SVM can be increased. By combining it with the other search techniques like PSO, GA etc. better feature selection can be done, as the SVM does not require the prior knowledge of data [3, 19]. However, many researchers found some limitations of SVM like: (1) The kernel selection for a problem (2) The functional speed of the machine in training and testing, (3) Slower Convergence rate at testing phase,(4) Choosing good quality kernel parameters, (5) Large requirements of memory space to implement the model (6) Choosing either parametric or non-parametric method for implementation[8].

III. PARTICLE SWARM OPTIMIZATION

Initially, particle swarm optimization (PSO) was developed for the continuous optimization problems. But for performing the feature selection, PSO was extended to compact with the binary data. There are different types of PSO are present in the literature. Figure 3 exhibits the simple process of particle swarm optimization method.

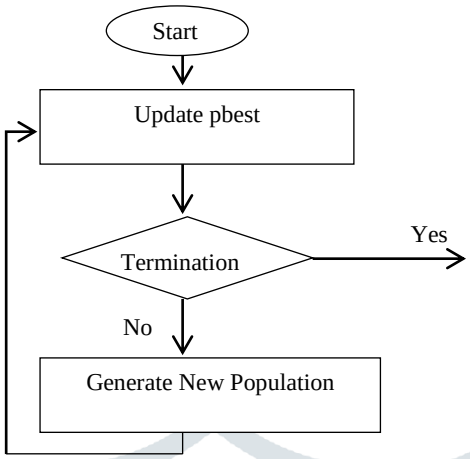

Figure 3. Flowchart for particle swarm optimization

Particle swarm optimization (PSO) is an evolutionary computational method, originally considered for multidimensional continuous space [2]. PSO and genetic algorithms (GA) are widely used optimization method for feature selection. Both the algorithms are populations based however PSO has some advantages over GA [mu, ne]. For example non-requirement of crossover and mutation operator makes it simpler than GA. It optimizes a problem by iterations, trying to improve a candidate solution in context of specified measures of quality. The problem is solved by having a population of candidate solutions, moving particles in search-space in accordance with simple mathematical formulae to improve particle's velocity and position. Sousa et al. designed two-stage PSO. In first stage three different PSO algorithms were implemented and tested against a Genetic Algorithm and Tree Induction Algorithm (J48). In the second stage the best classifier variants improved in terms of attribute type support and temporal complexity [22]. Li Yeh C. et al proposed an optimization algorithm named catfish binary particle swarm optimization (CBPSO. The quality of solution has measured by K-nearest neighbor method with leave-one-out cross-validation (LOOCV) [23].

As compared to other optimizations techniques, PSO has improved global optimization and lesser computing complexity. Sometimes, PSO based feature selection methods fall into local optimum and not able to provide optimal feature subset, particularly when there are large number of original features. Wang et al proposed a novel feature selection method, based on one-dimension real valued particle swarm optimization algorithm, for the selection of optimal feature subset [24]. A two-phase feature selection method using PSO, in which a core set of discriminatory features formed from the original feature space and then this set is successively extended by searching for additional discriminatory features. A nearest neighbor classifier is used to evaluate the performance of the proposed PSO feature selection method. There is a significant reduction of the feature space, in turn, an improvement in classification accuracy [25]. Particle swarm optimization is an optimization algorithm that can be used for feature selection as well as parameter optimization. A genetic algorithm (GA) approach to simultaneously optimize the SVM parameters and feature subset without degrading the SVM classification accuracy is suggested by Huang and Wang [26].

Zahran and Kanaan have used PSO for text categorization. In this study supervised learning algorithm Radial Basis Function (RBF) network is used as classifier which has a single hidden layer. In RBF, if the cluster centre is known, then the distance of the input pattern from the cluster can be measured. According to the author, it was the first research on Arabic data set that used PSO as feature selection. The performance of this algorithm is compared with document frequency and chi square. The experimental results show the superiority of the proposed algorithm than others [27]. Esmin et al proposed the generation method for fuzzy rule by learning from samples using PSO [28]. Because of its simplicity and easy implementation, PSO can usually applied in function optimization, neutral network training, machine study, the signal processing, model classification, automatic adaptation control etc. [29].

Several modifications have applied to the original PSO/ACO algorithm by Holden and Nicholas. These modifications used for extracting the classification rules and evaluated the new version of the PSO/ACO (i.e., PSO/ACO2) algorithm with two dissimilar particle fitness functions to illustrate the quality of performance of PSO/ACO2. The quality of the performance is significantly affected by these fitness functions [30]. Permana and Hashim designed Fuzzy Particle Swarm Optimization (FPSO) to enhance the speed of convergence and performance of fuzzy system [31].

Two filter feature selection techniques for classification are suggested by Liam et al. The first technique is based on binary particle swarm optimization (BPSO) method and mutual information of each pair of features that defines the significance and the redundancy of the particular feature subset. Second technique is based on BPSO and the entropy of every set of features, which estimates the significance and the redundancy of the selected feature subset. Decision tree is employed as the classifier to calculate the accuracy of classification [33]. Aghdam and Heidari have presented a feature selection technique based on PSO to

improve the performance of text categorization. According to the authors the complexity of this method is relatively low as they used simple classifier. The result of the experiment shows the superiority of the presented method. To show the superiority of the proposed method the authors compares it with GA, information gain and chi-square. PSO have the ability to converge quickly because of strong search capability, hence it can find minimal feature subset [4]. Shih-Wei Lin et al establishes the values of parameter while discovering a separation of features, without degrading the quality of classification. In this research, an approach is developed, called PSO+SVM for feature determination, is developed where radial basis function is used to obtain the results [33].

IV. COMPARATIVE STUDY

In this comparative study, a few research papers on feature selection are compared on the basis of classification and feature selection problems and represented in Table 1. Feature selection, used for high dimensional feature space and complex problems, gives better accuracy for SVM. Discrete particle swarm optimization algorithm for the feature subset selection problem in which an adaptive feature selection procedure which dynamically accounts for the relevance and dependence of the features, is compared with the tabu search and scatter search algorithms on publicly available datasets. The results show that the proposed discrete PSO algorithm is competitive in terms of both classification accuracy and computational performance. Combination of particle swarm optimization based feature selection algorithm and Radial Basis Function (RBF) networks as a text classifier used in text characterization and results show improvement in the performance of Arabic text categorization. In pattern recognition, a two-phase feature selection method that uses particle swarm optimization, in which in initial phase a core feature set is formed and then it is sequentially expanded by searching for additional discriminatory features. The nearest neighbor classifier is used for the performance evaluation of the proposed PSO feature selection method. There is substantial reduction of the feature space and improvement in classification accuracy. The feature selection process is effectively simplified by application of catfish effect on binary PSO. The K-nearest neighbor (K-NN) method with leave-one-out cross-validation (LOOCV) is used to evaluate the quality of the solutions. Cat Swarm Optimization Algorithm (CSO) is an optimization algorithm. Support Vector Machine (SVM), as a classifier, combined with Harmonious CSO (HCSO) algorithm, for feature selection, show better solution than CSO. Whole comparison indicates that various hybrid swarm optimization algorithms provide better classification accuracy than existing basic methods.

Table 1: Comparative Study

Title of paper	Algorithm(s) used	Parameter for Comparison	Research outlook
Feature selection of support vector machine based on harmonious cat swarm optimization (HCSO) [3]	SVM, HCSO	Classification Accuracy	• SVM is used as classifier. • HCSO has better experimental results than CSO for feature selection
Feature selection using particle swarm optimization in text categorization [4]	PSO, Nearest Neighbour Classifier (KNN)	Classification Accuracy	• Simple classifier 'Nearest Neighbour Classifier (KNN)' provides low complexity. • PSO have higher accuracy as compared to GA, information gain and CHI.
Improved binary particle swarm optimization using catfish effect for feature selection [23]	Catfish BPSO, KNN	Classification Accuracy	• Applied catfish effect to increase the performance of BPSO. • KNN+ leave-one-out cross-validation method is used to evaluate the quality of solutions
Identifying core sets of discriminatory features using particle swarm optimization [25]	PSO, KNN	Classification Accuracy	• Two phase feature selection with PSO forms initial core set of discriminatory feature from original feature subset. • The effectiveness depends upon the size of initial core set of features.
Text feature selection using particle swarm optimization algorithm [27]	PSO, Radial Basis Function (RBF) networks	Classification Accuracy	• It is the first research on Arabic dataset. • Radial basis function networks used as text classifier.
A discrete particle swarm optimization method for feature selection in binary classification problems [34]	Discrete PSO, Logistic Regression Model	Classification Accuracy	• Proposed method is compared with Tebu Search and scatter search algorithm. • Proposed Discrete PSO performed better for classification.
Feature selection for SVM [35]	SVM	Leave one out error	• This search performed better than wrapper method and filter method for feature selection • This method is applicable for multi-dimensional data.

V.CONCLUSION

In this paper we presented a review on particle swarm optimization and support vector machine algorithm. This paper examined that various hybrid swarm optimization algorithms provide better classification accuracy and computational results as compared to other feature selection methods.

REFERENCES

[1] Pritesh Vora, Bhavesh Oza. 2013. Improved data reduction technique in data mining. International Journal of Computer Science and Mobile Computing, 2:169-174.

[2] J. Jebamalar Tamilselvi, K.Sutha. 2015. A review of feature selection algorithms for data mining techniques. International Journal on Computer Science and Engineering, 7(6): 63-67.

[3] Kuan Cheng Lin, Kai Yuan Zhang, Jason C. Hung. 2014. Feature selection of support vector machine based on harmonious cat swarm optimization. International Conference on UBI-Media Computing and Workshops, 205-208.

[4] Mehdi Hossein Aghdam, Setareh Heidari. 2015. Feature selection using particle swarm optimization in text categorization, Journal of Artificial Intelligence and Soft Computing Research, 5 (4): 231-238.

[5] Divya Tomar, Sonali Aggarwal. 2014. A survey on pre-processing and post-processing techniques in data mining, 7(4):99-128.

[6] Poonam Bonde, Sachine Barahate. 2017. Classification of big data using support vector machine. International Journal of Emerging Technologies and Innovative Research, 4(3): 96-100.

[7] Kuan Cheng Lin, Kai Yuan Zhang, Yi Hung Huang, Jason C. Hung, Neil Yen. 2016. Feature selection based on an improved cat swarm optimization algorithm for big data classification. Journal of Supercomputing, 72(8):3210-3221.

[8] Janmenjoy Nayak, Bighanraj Naik, H.S. Behra. 2015. A comprehensive survey on support vector machine in data mining tasks: applications and challenges. International Journal of Database Theory and Applications, 8(1): 169-186.

[9] Sujun Hua, Zhirong Sun. 2001. Support vector machine approach for protein subcellular localization prediction. BIOINFORMATICS, 17(8): 721–728.

[10] A. Zien, G. Rätsch, S. Mika, B. Schölkopf, T. Lengauer, K. R. Müller. 2000. Engineering support vector machine kernels that recognize translation initiation sites. 16(9): 799-807.

[11] Yu Dong Cai, Pong Wong Ricardo , Chih Hung Jen , Kuo Chen Chou. 2004. Application of SVM to predict membrane protein types. Journal of Theoretical Biology, 226(6): 373–376.

[12] R. Burbidge, B. Buxton. 2001. An introduction to support vector machines for data mining. 3–15.

[13] Jin Hyung Kim, Bong Kee Sin. 2014. Online handwriting recognition. Handbook of Document Image Processing and Recognition Springer, 887-915.

[14] G. Y. Chen, W. F. Xie. 2007. Pattern recognition with SVM and dual-tree complex wavelets. Image and Vision Computing, 25(6): 960-966.

[15] Nahla Barakat, Andrew P. Bradley. 2007. Rule extraction from support vector machines: A sequential covering approach. IEEE Transactions on Knowledge and Data Engineering, 19(6): 729-741.

[16] Nahla Barakat, Andrew P. Bradley. 2010. Rule extraction from support vector machines: A review. Neurocomputing, 74(1–3): 178-190.

[17] Kunlun Li, Jing Xie , Xue Sun, Yinghui Ma, Hui Bai. 2011. Multi-class text categorization based on LDA and SVM. Procedia Engineering 15: 1963–1967.

[18] Abd. Samad Hasan Basari, Burairah Hussin, I. Gede Pramudya Ananta, Junta Zeniarja. 2012. Opinion mining of movie review using hybrid method of support vector machine and particle swarm optimization. Malaysian Technical Universities Conference on Engineering and Technology, 453-462.

[19] Chung Jui Tu, Li Teh Chuang, Jung Yang Chang. 2007. Feature selection using PSO-SVM. IAENG International Journal of Computer Science, 33(1).

[20] Mukesh, J. Vashasitha. 2017. Hybrid swarm optimization approaches: a survey. International Journal of Latest Trends in Engineering and Technology, 9(2):175-181.

[21] Neha, J. Vashishtha. 2016. Particle swarm optimization based feature selection. International Journal of Computer Applications. 146(6):11-17.

[22] Tiago Sousa, Arlindo Silva, Ana Neves .2004. Particle swarm based data mining algorithms for classification tasks. Parallel Computing 30(5-6):767–783.

[23] Li Yeh Chuang, Sheng Wei Tsai, Cheng Hong Yang. 2011. Improved binary particle swarm optimization using catfish effect for feature selection. Expert Systems with Applications, 38(10): 12699-12707.

[24] Jun Wang, Yan Zhao, Ping Liu. . 2010. Effective feature selection with particle swarm optimization based one dimensional searching. IEEE,978(1):4244-6044

[25] W. Pedrycz, B. J. Park, N. J. Pizzi. 2009. Identifying core sets of discriminatory features using particle swarm optimization. Expert Systems with Applications, 36(3): 4610-4616.

[26] Cheng Lung Huang, Cheih Jen Wang. 2006. A GA based feature selection and parameter optimization for support vector machines. Expert Systems with Applications, 31(2): 231-240.

[27] Bilal M. Zahran, Ghassan Kanaan. 2009. Text feature selection using particle swarm optimization algorithm. World Applied Science Journal (Special Issue of Computer and IT), 69-74.

[28] A. A. A Esmin. 2007. Generating fuzzy rules from the example by using the particle swarm optimization algorithm. International Conference on Hybrid Intelligent Systems (HIS 2007). 340-343.

[29] Dian Palupi Rini, Siti Mariyam Shamsuddin. 2011. Particle swarm optimization: techniques, system and challenges. International Journal of Computer Applications, 14(1): 19-27.

[30] Nicholas Holden, Alex A. Freitas. 2007. A hybrid PSO/ACO algorithm for classification. Conference Companion on Genetic and Evolutionary Computation, ACM Press, 2745-2750.

[31] Kurniawan Eka Permana, Siti Zaiton Mohd Hashim. 2010. Fuzzy membership function generation using particle swarm optimization. International Journal of Open Problems in Computer Science and Mathematics, 3(1): 1998-6262.

[32] Liam Cervante, Bing Xue, Mengjie Zhang, Lin Shang. 2012. Binary particle swarm optimization for feature selection: A filter based approach. IEEE Congress on Evolutionary Computation, 1-8.

[33] Shih Wei Lin, Kuo Ching Ying, Shih Chieh Chen, Zne Jung Lee. 2008. Particle swarm optimization for parameter determination and feature selection of support vector machines. Expert Systems with Applications, 35(4):1817-1824.

[34] Alper Unler, Alper Murat. 2010. A discrete particle swarm optimization method for feature selection in binary classification problems. European Journal of Operational Research, 206(3): 528-539.

[35] J. Weston, S. Mukherjee , O. Chapelle, M.Pontil, T.Poggio, V.Vapnik. 2001. Feature selection for SVMs. In NIPS 13.