

Formulation of Feature Selection with Support Vector Machine

Gend Lal Prajapati

Department of Computer Engineering
Institute of Engineering & Technology
Devi Ahilya University Indore-452017 INDIA

Arti Patle

Department of Computer Engineering
Institute of Engineering & Technology
Devi Ahilya University Indore-452017 INDIA

ABSTRACT

Basic question arises when classification came in picture classification accuracy, ensemble size, and computational complexity. Feature selection is importance for improvement and performance of classification algorithm. Classification algorithm may not scale up to the size of the full feature set either in sample or time but with feature selection help us to better understand the domain with Cheaper to collect a subset of predictors and Safer to collect a reduced subset of predictors. An important pre-processing step in classification tasks is called as, Feature selection its aims to minimize both the classification error rate and the number of features for inference knowledge in any domain. Feature selection is Minimum set F that achieves maximum classification performance of T (for a given set of classifiers and classification performance metrics). This paper proposes feature selection methodology which includes ranking, information gain and filter method concept. After the feature subset train SVM with RBF kernel for classification.

Keywords

Ranking; Feature; classification; kernel; filter; information.

1. INTRODUCTION

Feature selection is defined as “the search for a subset of the original measurement” Features are not the same thing as Information, Feature attempt to access information [2].

Feature \neq Information

Features are that provide optimal information between probability error and cost of accurate Classification. Today's world is full of information, in every object and entity many information is stored, Classification. Today's world is full of information, in every object and entity many information is stored, only thing is collection and summarization for inference knowledge [1]. Machine learning play an important role in this area because in ML include so many things like: classification, association, prediction, inference, preprocessing etc. in this research our evaluation for classification. If we want classify common everyday objects such as people, cars and animal, it's pretty clear that using the features of those objects to do the job. People and animals have legs, which are a feature that cars don't have. Cars have wheels that are a feature that people don't have [4]. By selecting the appropriate set of features, we can do a good job of classification. *Albert Einstein says “Make everything as simple as possible but not simpler”*. If we consider in machine learning than select the best feature for classification but optimal feature.

For the classification which is feature based we should have some knowledge about feature of object to make best prediction. From the knowledge of feature classification task is quit less complex [11]. For example, having wheels or not

distinguishes people from cars, but doesn't distinguish cars from trains. These are two different classification tasks. Depending on the classification task we are facing, different features or sets of features may be important, and knowing how we arrive at our knowledge of which features are useful to which task is essential.

Classification applies in many domains such as: medical, economy, Security, robotics etc. in recent era SVM classifier used in research field as well for the support of society, for example cancer diagnosis [6], intrusion detection, soil categorization, defect detection etc.

The classification process with supervised learning always involves two steps:

1.1 Training (with assessment) – this is where we discover what features are useful for classification by looking at many pre-classified examples.

1.2 Classification (with assessment) – this is where we look at new examples and assign them to classes based on the features we have learned about during training [12].

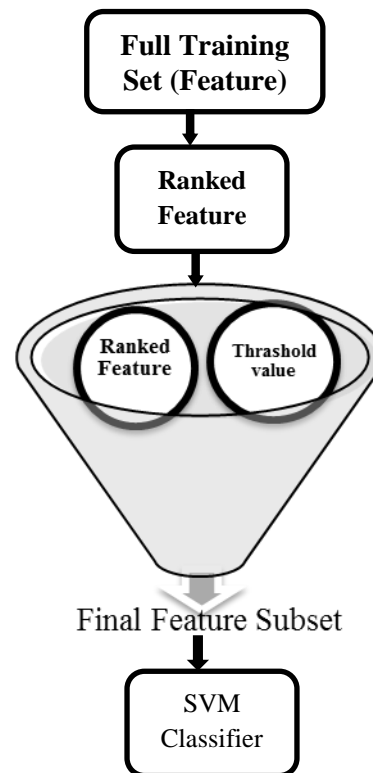


Fig 1: Feature Subset Selection

The task is to introduce a hypothesis (classifier) that accurately predicts the labels of novel instances. The learning of the classifier is inherently determined by the feature-values [1]. Theoretically high dimension of feature discriminating power, but in practical, with a limited amount of training data, excessive features will not only significantly slow down the learning process, but also cause the classifier to over-fit the training data as irrelevant or redundant features may confuse the learning algorithm for inference knowledge. Let F be a full set of features, F_i is a feature, and S_i (subset of feature) = $F - \{F_i\}$. These categories of feature relevance can be formalized [5].

Feature type in classification

Flat feature

- a. Filter method b. Wrapper Method
- b. Embedded Method
Streaming features, structured feature
- c. Group structure
Tree structure, Graph structure

2. SUPPORT VECTOR MACHINE

“Hype or Hallelujah?” is the provocative title used by Bennett & Campbell (2000) in an overview of Support Vector Machines (SVM). SVMs have best classification accuracy for that reasons SVM is currently a hot topic in the machine learning community [3]. SVM is one of the best supervised learning methodologies in data mining and machine learning field. Heart of the SVM is kernel function [1], in this formulation we consider RBF for classification.

2.1 SVM Construction

Given a training set of N data points $\{(x_i, y_i)\}$ $N_i=1$, with input data $x_i \in \mathbb{R}^n$ and corresponding binary class labels $y_i \in \{-1, +1\}$,

$$w^T \phi(x_i) + b \geq +1, \text{ if } y_i = +1$$

$$w^T \phi(x_i) + b \leq -1, \text{ if } y_i = -1$$

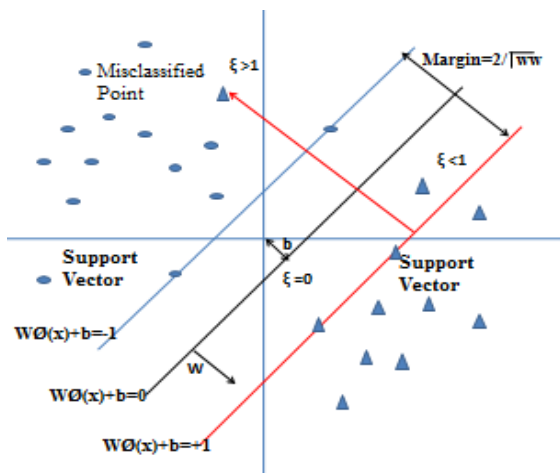


Fig 2: SVM Classification

The SVM classifier, according to Vapnik’s original formulation satisfies the following conditions:
Which is equivalent to:

$$y_i [w^T \phi(x_i) + b] \geq 1, i = 1, \dots, N$$

The non-linear function $\phi(\cdot)$ maps the input space to a high (possibly infinite) dimensional feature space [14]. In this feature space, the above inequalities basically construct a hyperplane $w^T \phi(x_i) + b = 0$ discriminating between both classes. In primal weight space the classifier then takes the form

$$y(x) = \text{sign}[w^T \phi(x_i) + b]$$

But, on the other hand, is never evaluated in this form. One defines the convex optimization problem:

$$\text{Min}_{w,b,\xi} \tau(w, b, \xi) = \frac{1}{2} W^T W + C \sum_{i=1}^n \xi_i$$

Subject to

$$y_i [w^T \phi(x_i) + b] \geq 1 - \xi_i, \quad i = 1, \dots, N$$

$$\xi_i \geq 0, \quad i = 1, \dots, N.$$

The variables ξ_i are slack variables which are needed in order to allow misclassifications in the set of inequalities (e.g. due to overlapping distributions). The first part of the objective function tries to maximize the margin between both classes in the feature space, whereas the second part minimizes the misclassification error. The positive real constant C should be considered as a tuning parameter in the algorithm [16].

The Lagrangian to the constraint optimization problem is given by

$$L(w, b, \xi; \alpha, v) = \tau(w, b, \xi) - \sum_{i=1}^n \alpha_i \{y_i [w^T \phi(x_i) + b] - 1 + \xi_i\} - \sum_{i=1}^n v_i \xi_i$$

The solution to the optimization problem is given by the saddle point of the Lagrangian, i.e. by minimizing $L(w, b, \xi; \alpha, v)$ with respect to w, b, ξ and maximizing it with respect to α and v . This leads to the following classifier:

$$y(x) = \text{sign} \left[\sum_{i=1}^n \alpha_i y_i K(x_i, x) + b \right]$$

Whereby $K(x_i, x) = \phi(x_i)^T \phi(x)$ is taken with a positive definite kernel satisfying the Mercer theorem. The Lagrange multipliers (α_i) are then determined by means of the following optimization problem (dual problem): $A = \pi r^2$

$$\text{Max}_{\alpha_i} - \frac{1}{2} \sum_{i,j=1}^n y_i y_j K(x_i, x_j) \alpha_i \alpha_j + \sum_{i=1}^n \alpha_i$$

Subject to

$$\sum_{i=1}^n \alpha_i y_i = 0$$

$$0 \leq \alpha_i \leq C, i = 1, \dots, N$$

The entire classifier construction problem now simplifies to a convex quadratic programming one does not have to calculate w nor in order to determine the decision surface [7]. Thus, no explicit construction of the nonlinear mapping $\phi(x)$ is needed. Instead, the kernel function K will be used.

Iris data set have tree type of classes which are differentiating with color. With the help of SVM, dataset should be classified. Classes of dataset should be classified with the classifier may be linear or non-linear.

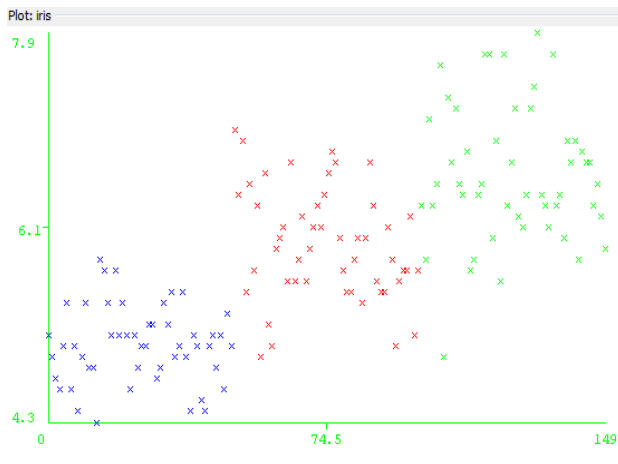


Fig 3: Visualization of Iris Data Set from UCI Repository in WEKA Tool

3. LITERATURE REVIEW

Optimal set of feature that achieves maximum classification performance for decision making or prediction in any domain is Feature selection [10]. This formulation combine with SVM classifier and it gives better result rather than full feature set.

Minh and Torre, discuss Features selection for support vector machine (SVM) classifiers are important for many of reasons such as generalization performance, computational efficiency, and feature interpretability. Traditional SVM classification with feature selection typically extracts features and optimizes SVM parameters independently [10].

This research enlightens supervised learning problems other classifiers unable to discard irrelevant features [8] (e.g. noise, outliers, redundant features) will affect the system performance which includes classification accuracy, computational efficiency, and learning convergence [9]. First, the implicit regularization achieved by feature pruning typically increases the generalization ability of classifiers; this generally leads to higher classification accuracy. Second, using irrelevant features also considerably increases the computation time. Third, too many features may render the convergence impossible, leading to random classification decisions. In addition to the system performance, identification of important variables that have intuitive physical interpretation is another critical requirement of many applications. Irrelevant features typically do not have intuitive justification [17]. Due to the aforementioned reasons, feature selection has been a central topic in a variety of fields including signal processing, computer vision, statistics, neural networks, pattern recognition, and machine learning.

With the feature selection scheme there are many strategies such as: Univariate or multivariate. In the univariate scheme, each feature is ranked independently of the feature space, while the multivariate scheme evaluates features in a batch way. Therefore, the multivariate scheme is naturally capable of handling redundant features. In the second step, the features with highest rankings are chosen to induce classification models [15]. In the past decade, a number of performance criteria have been proposed for filter-based feature selection such as Fisher score, methods based on mutual information and Relief and its variants. Up to extend mutual information used by many application classification tasks for relevancy.

Deng et al. in 2013 gives a feature-selection algorithm based on Support Vector Machine-Multiclass for hyperspectral visible spectral analysis. Issue of This this research paper is quality and safety of foods. Using high-precision spectral devices inspecting the food with technology trends by its high accuracy and nondestructive, in this the common obstacle is how to extract informative variables from rough data without losing valuable information. This paper proposes a novel feature selection algorithm Support Vector Machine-Multiclass Forward Feature Selection (SVM-MFFS). SVM-MFFS incorporate with wrapper and forward feature selection strategy, for emphasis the stability of spectral variables, and uses classical SVM as classification and regression model to select the most relevant wavelengths from hundreds of spectral data. This paper compares SVM-MFFS with Successive Projection Analysis and Uninformative Variable Elimination in the experiment of identifying different brands of sesame oil. This research results analysis is SVM-MFFS outperforms in accuracy, Receiver Operating Characteristic curve, Prediction and Cumulative Stability, and it will provide reliable results [16].

M. F. Akay in 2009 diagnosis Breast cancer from Support vector machines combined with feature selection. Research valuable if it beneficial for society, because of this purpose many researchers focus in many field like medical, economics, military, security, academics etc. in medical domain Breast cancer is the second largest cause of cancer deaths among women. At the same time, it is also among the most curable cancer types if it can be diagnosed early. Research efforts have reported with increasing confirmation that the support vector machines (SVM) have greater accurate diagnosis ability. In this paper diagnosis the breast cancer with SVM classification. In this research classification perform with feature selection for effective identification. Experiments have been conducted on different training-test partitions of the Wisconsin breast cancer dataset (WBCD), which is commonly used among researchers who use machine learning methods for breast cancer diagnosis. The performance of the method is evaluated using classification accuracy, sensitivity, specificity, positive and negative predictive values, receiver operating characteristic (ROC) curves and confusion matrix. The results show that the highest classification accuracy (99.51%) is obtained for the SVM model that contains five features, and this is very promising compared to the previously reported results. To evaluate the effectiveness method, conducted experiments on the WBCD. The importance of each feature is measured by F-score, and the SVM parameters are optimized by grid search [6].

Waske et al. shows the Sensitivity of Support Vector Machines to Random Feature Selection in Classification of Hyperspectral Data in 2010. In this paper supervised classification apply on hyperspectral data for land cover scenario. The accuracy of supervised land covers classifications effected by feature selection. For this purpose support vector machines (SVMs) and multiple classifier systems (MCS) used. MCSs based on SVM and random feature selection (RFS) are applied to explore the potential of a synergetic use of the two concepts. This paper investigated how the number of selected features and the size of the MCS influence classification accuracy using two hyperspectral data sets, from different environmental settings. In addition, experiments were conducted with a varying number of training samples. Accuracies are compared with regular SVM and random forests. Research Experimental results demonstrate that the generation of an SVM-based classifier

system with RFS significantly improves overall classification accuracy as well as producer's and user's accuracies [7].

The SVM is a universal learning machine for solving binary Classification. REMOTE sensing applications, such as land cover classification, provide a variety of important information for decision support and environmental monitoring systems.

Morariu, Vintan, and Tresp give the Feature Selection Methods for an Improved SVM Classifier in 2008. Today's era textual information is available online in vast amount. It is difficult to retrieval effective information from documents so the classification apply from, good indexing and summarization of document content. In recent years a growing number of categorization methods and machine learning techniques have been developed and applied in different contexts. Training with classifiers on large collections of documents, both the time and memory restrictions can be quite prohibitive. Justify both the issue application of feature selection methods to reduce the dimensionality of the document-representation vector. This research evaluated three feature selection methods: Random Selection, Information Gain (IG) and Support Vector Machine feature selection (called SVM_FS). These researches also investigate that the no of feature affect the classification accuracy. Whenever use low dimension accuracy increase but with large dimension accuracy affected [5].

Karabulut, et al. discusses a comparative study on the effect of feature selection on classification accuracy in 2012. This Research paper focus on Feature selection for interest to many research areas which deal with machine learning and data mining, because it provides the classifiers to be fast, cost-effective, and more accurate. This paper focus on the effect of feature selection on the accuracy of Naïve Bayes, Artificial Neural Network as Multilayer Perceptron, and J48 decision tree classifiers is presented. These classifiers are compared with fifteen real datasets which are pre-processed with feature selection methods. Up to 15.55% improvement in classification accuracy is observed, and Multilayer Perceptron appears to be the most sensitive classifier to feature selection. In this study we have investigated the influence of feature selection on three classifiers Naïve Bayes, MLP and decision tree J48 using fifteen real-life datasets. We observed that MLP is the most affected classifier; ten of the used datasets are more accurately classified by preprocessing of at least one feature selector. The classification accuracy is improved up to 15.55% in Post-operative dataset. It is also observed for Naïve Bayes classifier the Gain Ratio, for MLP the Chi-square and for J48 the Information Gain is the most positively effective feature selection algorithm [13].

Anthony and Ruther Compare Feature Selection Techniques for SVM Classification in 2011. Environmental monitoring, modeling, map making and revision and urban studies all applications are feasible by classification. These entire fields have information, only thing is extraction of information for this SVM used for classification and with feature selection it is more effective It involves selecting a subset from the original set of features (e.g. bands) that captures the relevant properties of the data (Gilad-Bachrach et al., 2004) to enable adequate classification (Wu and Linders, 2000). Researchers used Exhaustive Search and Population Based Incremental Learning for appropriate feature selection techniques with Support Vector Machine classification and so is Thornton's separability index an appropriate criterion function.

Thornton's separability index=

$$\left[\sum_{i=0}^n (f(x_i) + f(x_i') + 1) \bmod 2 \right] / n$$

Where x is feature, x' is the nearest neighbor of x, n is the number of points.

Paper follow the approach is called the 'one against one' approach and involves constructing a machine for each pair of classes resulting in N(N-1)/2 machines. When applied to a test point, each classification gives one vote to the winning class and the point is labeled with the class having most votes. The second approach involves the 'one against all' approach where by the N class dataset is divided into N two-class cases. Proponents of the 'one against one' approach contend that it has the advantage of avoiding highly unbalanced training data.

4. TRAINING SUPPORT VECTOR MACHINE WITH FEATURE SELECTION: METHODOLOGY

Motivation of Feature Selection technique is increase classification accuracy and performance. In the review of literature various feature selection method are given by researcher. As Feature selection can remove the irrelevant features and improve the performance of learning systems, it is a crucial step in machine learning [13]. This research paper proposes the efficient feature selection method which is helpful for performance increment to train SVM. Feature generation approach has many types some as:

4.1 Feature Selection Generation

4.1.1 Forward Successors Generation

- a. Starting with $X' = \emptyset$.
- b. Adds features to the current solution X' , among those that have not been selected yet.
- c. In each step, the feature that makes J be greater is added to the solution. Operator is $O(n)$.

$$X' = X' \cup \{x_i \in X - X' | J(X' \cup \{x_i\}) \text{ is bigger} \}$$

4.1.2 Backward Successors Generation

- a. Starting with $X' = X$.
- b. Removes features from the current solution X' , among those that have not been removed yet.
- c. In each step, the feature that makes J be greater is removed from the solution. Operator is $O(n)$.

$$X' = X' - \{x_i \in X' | J(X' - \{x_i\}) \text{ is bigger} \}$$

4.1.3 *Correlation Coefficient:* These measures quantify how strongly two features are associated with one another is a classical measure that still use for these methods.

$$r = \frac{\sum_i (x_i - \bar{x}_i)(y_i - \bar{y}_i)}{\sqrt{\sum_i (x_i - \bar{x}_i)^2} \sqrt{\sum_i (y_i - \bar{y}_i)^2}}$$

Table 1. Dataset Description with Selected Feature with Filter & Wrapper Method

Relation Name	No. of Instance	No. of Attribute	No. of Selected Attribute With Filter Method	#No. of Selected Attribute With Wrapper Method
Audiology	226	70	6	3
Car	1728	7	1	1
Dermatology	366	35	16	1
liver-disorders	345	7	1	1
Sick	3772	30	1	1

This table shows the selected feature from wrapper and filter method. Feature selections method like filter and wrapper follow feature generation methods [18]. In this paper use the filter method for final feature selection. These are the dataset in which feature are selected with filter and wrapper method. In this table dataset car, dermatology etc. have only one selected feature that is not sufficient for accurate classification. But in case of audiology dataset selected feature is sufficient for classification with SVM. These are the example set for classification from which machine can be trained for testing of new experiment.

4.2 Proposed Methodology

Research hypothesis wants to find out $f = (f_1, \dots, f_N)$ from full training set of features $F = (F_1, \dots, F_N)$. So we focus on ranking and then filter method for subset selection. We propose a methodology for feature selection and classification with SVM.

Algorithm shows the procedure for feature selection and classification with SVM including RBF kernel. Algorithm has input with training set with class variable.

Algorithm: Formulation of Feature Selection with SVM

Require:

Input: Training Sample, $X_0 = [X_1, X_2, \dots, X_k]$,
Class labels, $Y = [Y_1, Y_2, \dots, Y_k]$
Output: Classification accuracy.

- 1: Initialize: Subset of Input Space features $S = [1, 2, \dots, n]$.
- 2: Calculate Feature ranked list with information gain $R = []$, until $S = \emptyset$.
- 3: Select threshold value n ; reject the feature whose feature rank is less than value n .
- 4: Update feature ranked list R with some threshold value than modified rank list is R_m .
- 5: Apply Filter subset selection on $R_m = []$.
- 6: Using filter method obtain Rank List is R_f
- 7: Train SVM classifier with RBF kernel function.

4.3 Example Run

For the example view take Labor negotiation dataset from UCI machine learning library. F-score for feature ranking: Instances x_i , $i = 1, \dots, n$ the F-score of the j^{th} feature is defined as:

$$F(j) = \frac{(\bar{x}_j^{(+)} - \bar{x}_j)^2 + (\bar{x}_j^{(-)} - \bar{x}_j)^2}{\frac{1}{n_+ - 1} \sum_{i=1}^{n_+} (x_{i,j}^{(+)} - \bar{x}_j^{(+)})^2 + \frac{1}{n_- - 1} \sum_{i=1}^{n_-} (x_{i,j}^{(-)} - \bar{x}_j^{(-)})^2}$$

Here n_+ and n_- are the number of positive and negative instances, respectively; \bar{x}_j , $\bar{x}_j^{(+)}$, $\bar{x}_j^{(-)}$ are the average of the j^{th} feature of the whole, positive-labeled, and negative-labeled data sets; $x_{ij}^{(+)}/x_{ij}^{(-)}$ is the j^{th} feature of the i^{th} positive/negative instance.

Some dataset are processed in WEKA tool for ranking and information gain according to algorithm. These selected features are executing with filter method and apply SVM classifier.

Following are the feature ranking of Labor-neg-data dataset [20].

Relation: labor-neg-data
Instances: 57
Attributes: 17
Ranked Features R:

- 0.2948 2 wage-increase-first-year
- 0.1893 3 wage-increase-second-year
- 0.1624 11 statutory-holidays
- 0.1341 14 contribution-to-dental-plan
- 0.1164 16 contribution-to-health-plan
- 0.1091 12 vacation
- 0.0855 13 longterm-disability-assistance
- 0.0717 9 shift-differential
- 0.0548 7 pension
- 0.0484 5 cost-of-living-adjustment
- 0.0333 15 bereavement-assistance
- 0.0307 4 wage-increase-third-year
- 0.024 10 education-allowance
- 0.0195 8 standby-pay
- 0 1 duration
- 0 6 working-hours

Classification Accuracy with SVM (without using any feature selection method) = 64.91%

Selection of feature, Setting threshold value $n = 0.5$, large value of threshold gives less no of feature from that classifier have not suitable accuracy. With small value of threshold gives large number of feature so that classification accuracy affected, for that reason take average value of threshold [19].

From consider threshold value, feature ranked list R_m is:

- 0.2948 2 wage-increase-first-year
- 0.1893 3 wage-increase-second-year
- 0.1624 11 statutory-holidays
- 0.1341 14 contribution-to-dental-plan
- 0.1164 16 contribution-to-health-plan
- 0.1091 12 vacation
- 0.0855 13 longterm-disability-assistance
- 0.0717 9 shift-differential
- 0.0548 7 pension

Apply filter method for feature selection on R_m :
Then Feature Rank R_f is

Feature No. - 2, 3, 11, 13.
2 Features: wage-increase-first-year
3 wage-increase-second-year
11 statutory-holidays
13 longterm-disability-assistance

Now the Classification Accuracy with R_f using SVM:
71.56%

5. CONCLUSION

Training and learning of the SVM Classifier determined by the feature space of training set. Research paper formalizes a feature selection method with ranking, information gain and filter for feature subset of dataset. This approach is demonstrated with an example. As after apply the method the classification accuracy is 71.56% instead of 64.9% (without using feature selection method). SVM is a representative of discriminative learning—i.e. with all corresponding advantages (power) and limitation. With the feature selection overcome the limitation of SVM with dataset length and reduction of dimension. We also investigated whether always selected feature is appropriate for classification. Fundamentally the proposed method is utilized for effective classification using feature selection with SVM.

6. REFERENCES

- [1] H. Liu and H. Motoda. 1998. "Feature Selection for Knowledge Discovery and Data Mining", Kluwer Academic, Boston.
- [2] J. X. Dong, A. Krzyzak and C. Suen. 1998. "Fast SVM training algorithm with decomposition on very large datasets", IEEE Transaction Pattern Analysis Mach. Intell., Vol. 27 (4), pp. 603–618.
- [3] Y. Liu and F. Zheng. 2006. "A novel feature selection method for support vector machines", Pattern Recognition Society, Published by Elsevier Ltd, vol. 39, pp. 1333-1345.
- [4] H. Liu, Zhao and Zheng. 2007. "Searching for Interacting Features", IJCAI, vol. 7, pp. 1156-1161.
- [5] D. Morariu, L. Vintan, and V. Tresp. 2008. "Feature Selection Methods for an Improved SVM Classifier", World Academy Science, Engineering and Technology, vol. 2, 2008.
- [6] M. Fatih Akay. 2009. Support vector machines combined with feature selection for Breast cancer diagnosis, Expert Systems with Applications An International Journal, Volume 36 (2), pp. 3240-3247.
- [7] B. Waske, S. Linden and J. Benediktsson. 2010. "Sensitivity of Support Vector Machines to Random Feature Selection in Classification of Hyperspectral Data", IEEE Transactions on Geoscience And Remote Sensing, Vol. 48 (7), pp. 2880 – 2889.
- [8] S. Ding and Li Chen. 2010. "Intelligent Optimization Methods for High-Dimensional Data Classification for Support Vector Machines", Intelligent Information Management, vol. 2, pp. 159-169.
- [9] G. L. Prajapati and A. Patle. 2010. "On Performing Classification Using SVM with Radial Basis and Polynomial Kernel Functions", ICETET-2010 3rd International IEEE Conference, pp. 512 – 515.
- [10] H. Minh and F. Torre. 2010. Optimal Feature Selection for Support Vector Machines, Robotics Institute, Carnegie Mellon University USA, vol. 43(3), pp. 584-591.
- [11] G. Wang and Liu. 2011. "A Hybrid Feature Selection Method for Data Sets of thousands of Variables", Advance Computer Control (ICACC), vol.2, pp. 288-291.
- [12] Q. Liu, C. Chen and Y. Zhang. 2011. "Feature selection for support vector machines with RBF kernel", Artificial Intelligent Review, vol. 36, pp. 99–115.
- [13] E. Karabulut, S. Özel and T. İbrikçi. 2012. "A Comparative Study on The Effect of Feature Selection on Classification Accuracy", First World Conference on Innovation and Computer Sciences, Vol. 1, pp. 323–327.
- [14] Y. Jeong, I. Kang and M. Jeong. 2012. "A New Feature Selection Method for One-Class Classification Problems", IEEE TRANSACTIONS on Systems, Man and Cybernetics—part C: Applications And Reviews, Vol. 42, No. 6, pp. 103-114.
- [15] J. Tang, S. Alelyani and H. Liu, "Feature Selection for Classification: A Review" <http://core.ac.uk/download/pdf/23799612.pdf>, accessed on: 06 May 2015.
- [16] S. Deng, Y. Xu and L. Xiaoli. 2013. "A Feature-Selection Algorithm Based On Support Vector Machine-Multiclass For Hyperspectral Visible Spectral Analysis", Journal of Food Engineering, Vol. 119 (1), pp. 159–166.
- [17] G. Wang, Song, Qinbao and J. Ni. 2013. "A Fast Clustering-Based Feature Subset Selection Algorithm for High-Dimensional Data", Knowledge and Data Engineering, IEEE Transactions, vol. 25, pp. 1-14.
- [18] G. L. Prajapati and A. Patle. 2014. "Effective Classification with Dimension Reduction on Support Vector Machine", International Journal of Science, Technology and Management, Vol. 6, pp. 1-13.
- [19] P. Ghamisi. 2015. "A Novel Feature Selection Approach Based on FODPSO and SVM", IEEE Transactions on Geoscience and Remote Sensing, vol. 53(5), pp. 2935 – 2947.
- [20] Machine Learning Library, center for Machine Learning and Intelligent System, <http://archive.ics.uci.edu/ml/> accessed on: April 2015.