International Journal of Pure and Applied Mathematics
Volume 117 No. 21 2017, 891-910
ISSN: 1311-8080 (printed version); ISSN: 1314-3395 (on-line version)
url: http://www.ijpam.eu
Special Issue

ijpam.eu

RMCFSVM-ABC: A Regularized Monotonic Cascaded Fuzzy Support Vector Machine (RM) with Artificial Bee Colony (ABC) Optimization for Data Mining with Prior Knowledge

[1]A. Shanmugapriya and [2]N. Tajunisha

[1]Head of Department-Computer Science,

Dr.R.V Arts & Science College.

tarunika2015@gmail.com

[2]Department of Computer Science,

Sri Ramakrishna College of Arts & Science for Women,

Coimbatore.

Abstract

In recent days, researchers have tried to integrate the prior knowledge with data mining concepts, which seems to be an interesting technology, but it is a great dispute. This idea should be achieved to attain the targets of many research work. Previously different research work has been suggested for gaining knowledge. Earlier research work of new Map reduced based Regularized Monotonic Fuzzy SVM (MRRMC-FSVM) model deals with this problem. Nevertheless, the Cascade MRRMC-FSVM can be distributed over various processors with modified map reduced framework overhead and demands less memory, resolve the scalability issue, because the kernel matrices were little compared with the regular MRRMC-FSVM. Need to choose the virtual pairs at any order, which will be defined according to the less and high random subsets of the input dataset. But, overall monotonicity isn’t given back exactly, this is a demerit, though the output of the experiment didn’t give any desirable variations for a monotonicity constraints. So, in the proposed work, these issues were resolved through the optimal framework such as Hybrid MRRMC-FSVM with Artificial Bee Colony (ABC) (HMRRMC-FSVM-ABC). Here, instead of the random selection, it will automatically choose the virtual pair which will be defined according to the less and high random subsets of the input dataset. The automatic selection of the virtual pairs was done according to the ABC procedure with minimum and maximum of certain random subsets of the input dataset, this value is said to the fitness value of ABC.

This will exactly provide the entire monotonicity and the experiments provide important variations for a greater number of monotonicity constraints. The proposed work's judgement is done in the matlab simulation environment from which it is confirmed that the proposed research method tends to give a good result when compared to the current approaches, mainly with respect to the accurate knowledge learning.

Key Words: Monotonicity, optimal virtual pair selection, knowledge learning, random subsets.

1. Introduction

An overwhelming progress has been done in Data mining concepts in these last ten years of research work [1]. But still there occurs a gap among the output of data mining concepts and the steps proceeded for them. So human intervention should fill this gap, but it will reduce the effectiveness of the process and its scope [2]. For instance: through data mining concept, it discloses that a purchase of diapers at a supermarket is frequently accompanied with the beer, but we can speculate that the buyer is probably a new father, and suggest other products suitable for this statistics. While doing the decision-making process, it is required to consider a chain of inferences, and, moreover we have numerous data which is relevant to few inference steps that permit us to automate them by learning from the data, we frequently have no data for most of the other steps, so this leads to a labour-intensive bottleneck [3].

By shrinking this knowledge gap" will increase the advantages which highly outstrip the merits achievable from the enhancing the process which are previously effective. Data pre-processing [4] is another area which lack in portable, machine-understandable knowledge has a great cost. In data mining process, it is truth that most the cost is spent in data gathering, integration, cleaning, transformation, etc.; commonly in initiating the issue, which can be resolved through data mining tools (classification, clustering, etc.). Regardless to this, development in this area seems to be inadequate than the development in cleansing the modelling tools additionally. Normally, this tool is general, creating them more attractive goal for the research work, whereas the pre-processing tends to be very domain-specific [5].

As a consequence, in data mining project, human knowledge were integrated with pre-processing , but they are enciphered in opaque scripts and SQL commands, and in the next project it has be begin from the scratch [6]. Attractively, this knowledge should be enciphered in a translucent and modular format which can be reused quickly. In the modeling phase, here the levels were advanced, the utilization of their ideas (Knowledge) stays as an obstacles. The efforts of data mining concepts were hardly successful in the 1st try; a cycle of mining, investigating the output and re-doing the mining with an enhanced model is needed.

Target of this cycle is to integrate the human ideas into the model. This is an repeated process, since it will be simple for the human being to up knowledge "on demand," in response to the output and failures of the data mining system, instead of giving required knowledge prior [7]. Different tools are available from statistical and computational view; simple interface will not be available for the humans to proceed. Generally, knowledge is essentially integrated by working with various data mining algorithms, variations of them, parameter settings, combinations of techniques, etc. so, this result in demanding data mining experts with advanced degrees that restricts its dispersions. This permits the users (human) to give their ideas absolutely. For instance: by validating the result and giving justification, which will enhance easy utilization, productivity, and reusability [8].

Consider the example of loan-approval case [9], which is explained previously, in data classification applications, to some extent the person will have prior knowledge, so an increase in an input variable (attribute) should not head to the minimization of the class labels. Other domains example stats that, this kind of knowledge avail in medicine (e.g., smoking raises the likelihood of vascular diseases) and economics (e.g., house prices maximization or minimization depends on the location). Some helpful details regarding the central issue along with the training data, will be given by the prior knowledge of monotonicity which is obtained from domain experts, past practical experience, and the literature. Some monotonicity constraints has to be included in the classification model as like FSVM, while considering the prior knowledge regarding the data. Knowledge can be drew-out from the monotonicity constraints, when it is integrated with the classification technique, this knowledge will be more reasonable and comprehensible [11].

In our work, the automatic selections of the virtual pairs were done according the artificial bee colony (ABC) with minimum and maximum of specific random subsets of the input dataset is assumed as the fitness value of ABC. This will ponder the entire monotonicity and experimental output provides important variations for the higher monotonicity constraints.

The implemented proposed methodology is given as follows: 1st section: prior knowledge's general details were given. 2nd section comprises of: related work details were given. 3rd section comprises of: discussion regarding the proposed research method for optimal and reliable learning of the proposed research methodology. 4th section comprises of proposed methodologies experimental output with respect to the performance comparison measures. 5th section comprises of: conclusion of the research work is provided according to the acquired output.

2. Related Works

We have two main approaches which take over the issues which has prior knowledge of monotonic properties, in data-mining literatures about monotonicity constraints.

One method is enforcing the relabeling technique for those data missing monotonicity [12]. Second one is to include the monotonicity constraints straight-away to the optimization modeling settings [13]. In order to apply the monotonicity, Evgeniou, Boussios and Zcharia [14] and Doumpos and Zopounidis [15] simulated a heap of monotonic data to develop the monotonicity constraints, which is done in the latter method. So this leads to the raise difficulty of the issue computation-wise.

A LSSVM regression model with monotonicity constraints is established by Pelckmans et al. [16]. Here, rather than utilizing the simulated data, the input data is utilized to develop the monotonicity constraints, and also they consider the input data follow a linear order and the bias term is eliminated. Nevertheless, those considerations can't be enforced in the real-time applications. Furthermore, in this model, the sparseness is gone. SO a new SVR model with monotonicity constraints is proposed, in order to proceed with this shortcomings happened in the previously addresses studies, this new model has inequalities and is depends on the partial order in the input data.

For classification issues with ordinal attributes, the class attribute should raise every explaining attributes or few of them; this is termed as classification problems with the monotonicity constraints [17]. This issue is normally faced in real-life applications like bankruptcy risk prediction [18], finance [19], breast cancer diagnosis [20], house pricing [21], credit rating [22] and many others.

The significance of the classification with monotonicity constraints is addressed in [23], which provides a work to enquire the possibility of the monotonicity constraints to bias machine learning systems help to learn the exact and meaningful one. M. Doumpos, C. Zopounidis [24] suggested a monotonic support vector machines for rating the credit risk. It utilizes the monotonicity "hints" to generate the virtual samples to enforce the monotonic conditions in order to indicate the special prior domain knowledge about the issue. Greek industrial firm's huge samples of the experimental output explain that the monotonicity condition minimizes the risk of over fitting, hence it proceeds to the models with greater capability of the prediction. S. Wang enforced the neural network with monotonicity property as a non-parametric efficiency analysis method in order to get the knowledge about the effective analysis for the private and public organizations [25].

3. Optimal Virtual Pair Selection Based Priori Knowledge Learning

The aim of our work is to incorporate the study is to integrate monotonic earlier knowledge in HMRRMC-FSVM-ABC for leading the learning procedure, by including the inequality constraints, which is acquired from the experts' knowledge about the target function. So, the formal explanation about the problem statement is provided in this section. Initially, the idea of monotonicity is elaborated and then explanation of the assisting ability of HMRRMC-FSVM-ABC in learning process is given. Next, the monotonically constrained FSVM (MC-FSVM) classification model is deduced. Third point is, the scalability issue of RMC-FSVM formulation and its monotonic property of the MR framework is explained theoretically. Finally, with few samples it explains the ways were given for the monotonicity constraints to help in enhancing the learning performance. Our work, automatically learn the selection process of virtual pairs which are defined according to few minimum and maximum of specific random subsets of the input dataset. Based on artificial bee colony (ABC), this selection process is done; the input dataset is assumed to be the fitness value of ABC.

Monotonic Constraints

In a modelling problem or project, there occurs a frequent state, that the functional form of an acceptable model is restricted; this is because of business considerations or due to the variety of scientific question which is enquired. A very strong belief there in few cases, where the true relationship has some quality, restriction can be utilized to enhance the predictive performance of the model.

A common variety of constraint in this conditions is that the specific features bear a monotonic relationship to the predicted response:

$$f(x_1,x_2,\ldots,x,\ldots,x_{n-1},x_n)\leq f(x_1,x_2,\ldots,x',\ldots,x_{n-1},x_n) \quad (1)$$

Whenever $x \leq x'$ $x \leq x'$ is an increasing constraint; or

$$f(x_1,x_2,\ldots,x,\ldots,x_{n-1},x_n) \geq f(x_1,x_2,\ldots,x',\ldots,x_{n-1},x_n)$$
$$f(x_1,x_2,\ldots,x,\ldots,x_{n-1},x_n) \geq f(x_1,x_2,\ldots,x',\ldots,x_{n-1},x_n) \quad (2)$$

Whenever $x \leq x'$ $x \leq x'$ is a decreasing constraint.

Monotonically Constrained Fuzzy Support Vector Machine Model

MC-FSVM, is a classifier model, which is constructed by including the monotonicity constraints, also it depends on the provided virtual examples, to the traditional FSVM model. Model derivation is provided only for the nonlinear cases, where for the linear model is derived by altering the kernel with the identity function. In many real world classification issues, few training points play a vital role that the others, FSVMs work on this fact. This significant point has to be classified perfectly, but less concern is given for the less significant points which are misclassified. A fuzzy membership function is determined so that every training point $\mathbf{x}_k$ is allotted with a value s_k , which is considered to the significant, $\mathbf{x}_k$ is classified as one class, and the value of $(1 - s_i)$ the degree to which misclassification is meaningless. For a set of observed data $\{(\mathbf{x}_k , y_k, s_k) | k = 1, 2, . . ., N\}$, the primal FSVM model [26] can be indicated as

$$\min J\,(w,e) = \frac{1}{2}w^T w + C\sum_{k=1}^{N} s_k e_k \quad (3)$$

Subject to $y_k(w^T\varphi(x_k) + b) \geq 1 - e_k, k = 1,2, \ldots, N, e_k \geq 0, k = 1,2, \ldots, N$

The function $\varphi(x) : R^n \rightarrow R^{nh}$, $n_h > n$, convert the input data into a higher dimensional feature space, where the classification issues are separable, or at least the segregation opportunity is high. Note that in (3), when $\varphi(x)$ is the identity function, then it is perfectly the FSVM, which is utilized for linearly separable issues. For the FSVM model in (3), the target function (or the classifier) has the form $y(x) = \text{sign}\,[w^T \varphi(x) + b]$. If it is monotonic, without loss of generality, the monotonicity can be signified as:

$$w^T\varphi\left(\underbrace{x}\right) \leq w^T\varphi(\tilde{x}), \quad \text{For observation } \underbrace{x} \leq \tilde{x} \quad (4)$$

A group of virtual examples or noticed pairs of data points are provided; one creates the monotonicity constraints according to the addressed inequality. For notational accessibility, we represent the group of observed monotonic constraints as MC, and

$$MC = \{(\underbrace{x}_i, \check{x}_i) | \text{ for all observed } \underbrace{x}_i, \check{x}_i, i = 1, \ldots, M\} \quad (5)$$

Where each pair $(\underbrace{x}_i, \check{x}_i)$ represents the corresponding constraint $w^T \phi(\underbrace{x}_i) \leq w^T \phi(\check{x}_i)$.

Cascaded FSVM using Map Reduce Framework

A distributed architecture is established to advance the FSVM implementations; here the smaller enhancement issues were resolved separately and distributed over various processors. At the same time it is assured to converge the globally optimal solution. To speed up the FSVMs, need to remove the non-support vectors as early as possible from the optimization process, which is confirmed to be an efficient strategy. MapReduce [27] is considerably varied from the model which is examined earlier to the parallel computation, since it interleaves parallel and sequential computation.

MapReduce Basics: In this paradigm, the basic unit of details assumes the details regarding the about support vector SV_i and their data are binary strings. Set of $\langle SV_i, \text{data}\rangle$pairs is the input of the MapReduce algorithm. Removing the non-support vectors happens in three stages: the map stage, the shuffle stage and the reduce stage.

In the map stage, the mapper μ considered as input for a set $\langle SV_i, \text{data}\rangle$pairs, and generates a set of new $\langle SV_i, \text{data}\rangle$pairs as output. The critical thing here is that, map operation works with only one pair at a time. Parallelization becomes easy here, as various $\langle SV_i, \text{data}\rangle$ for the map can be processed by various machines. In shuffle stage, the basic system executes the Map Reduce and transmits entire $\langle SV_i, \text{data}\rangle$values which are correlated with a single $\langle SV, \text{data}\rangle$ to the same machine. In reduce stage, reducer ρ considers entire value of $\langle SV, \text{data}\rangle$ correlated with the individual key k, and outputs a multiset of $\langle SV, \text{data}\rangle$pairs with the same key, k. This features the subsequent feature of MapReduce computation: maps have to be completed before initiating the reduce stage; because the reducer can use the entire values with the similar key, sequential computations can be done on these values. In the reduce step, the parallelism is used by noticing the reducers work on various keys, which can be implemented at the same time. On the whole, a MapReduce paradigm has different rounds of various map and minimize the functions, done one after the other. A map reduce program has the series $\{\mu_1, \rho_1, \mu_2, \rho_2, \ldots, \mu_R, \rho_R\}$ of mappers and reducers. The input is a multiset of $\langle SV_i, data\rangle$pairs represented by U_0. To implement the program on input U_0: For $r = 1, 2, \ldots, R$, do:

- Execute Map: Feed every pair U_0 in U_{r-1} to mapper μr, and continue. The mapper will produce a sequence of support vectors. Let U_r' be the multiset of key $\langle SV_i, data\rangle$ pairs output by μ_r.
- Shuffle: For every k, let $V_{k,r}$ be the multiset of values vi such that $\langle SV_i, data\rangle \in U_r'$. The underlying MapReduce implementation built the multisets $V_{k,r}$ from U_r' .
- Execute Reduce: For every k, feed k and some arbitrary permutation of $V_{k,r}$ to a separate instance of reducer ρ_r, and run it. The reducer will produce a sequence of tuples $\langle k, data_1'\rangle, \ldots \langle k, data_n'\rangle$, ... Let Ur be the multiset of $\langle SV_i, data\rangle$ pairs output by ρ_r , that is, $U_r = \bigcup_K \rho_r\ (k;\ data_{k,r})$. When combining the two results of SVM_1 and SVM_2, can initialize the enhancement of SVM_3 to various starting points. For a good performance put an effort to advance the optimization as much as feasible in every stage. This depends on how the data are separated basically, how partial results are combined and how well and optimization can begin from the partial results given by the earlier stage. This method assures the convergence to the global optimum.

Artificial Bee Colony (ABC) Algorithm

Artificial Bee Colony (ABC) helps in choosing the automatic virtual pair. This method guarantees the reflection the entire monotonicity and good experimental result is acquired in classification task and Convergence to the global optimum and high accuracy deficiency.

Behaviour of Real Bees

According to the reaction–diffusion equations [28], foraging behaviour of a honeybee colony is modelled by Tereshko. This model, heads to the collective intelligence of honeybee swarms and also it has three important components: food sources, employed foragers, and unemployed foragers, and it determines the two superior modes of the honeybee colony behaviour: hiring food source and abandonment of a source. Tereshko analyze the primary components of his model as below:

1. Food Sources: A forager bee judge various properties which is based on the food sources such as its closeness to the hive, richness of the energy, taste of its nectar, and the ease or complexity in drewing-out this energy, to choose the food source. For the simplicity, the quality of a food

source can be denoted through the quantity, though it is related to different parameters which are addresses above.

2. Employed foragers: This is hired, at a particular food source, where she is presently used. She will transfer the carried information regarding the specific food source with other bees in the hive. The information is about the distance, the direction and the profitability of the food source.
3. Unemployed foragers: A forager bee which waits for the food source to be used is termed as unemployed. IT may be scout who look for an environment randomly or an onlooker, which looks to identify the food source by means of the information provided by the employed bee. The scouts count is about 5–10%.

In creating collective knowledge, interchanging the details regarding the food source between the bees is the significant step. Analyzing hive feasibly differentiates few parts which are most commonly occurring in the entire hives. Dancing area is considered to a significant part of hive in terms on interchanging details. In this area, interaction between the bees about the quality of food sources happens. waggle dance is nothing but a related dance, because details regarding the entire rich sources will be accessible to an onlooker on the dance floor, apparently , enormous amount of dance will be watched and most profitable source will be chosen. Higher probability is there to choose the profitable source, since more information will be flowing regarding the profitable sources. Employed foragers share their information with the food source, which has a probability that is highly proportional, and sharing through waggle dancing takes much time. So, hiring process is proportional to profitability of a food source.

In this algorithm, food source's position points to the feasible output for the optimization issue and the nectar amount of a food source with respect to the quality (fitness) of the correlated solution. The employed bees or onlooker bees' count is equal to the amount of output in the population. In 1st step, ABC produces a randomly distributed initial population P (C = 0) of SN solutions (food source positions), where SN indicates the employed bees or onlooker bees size. Every solution x_i ($i = 1, 2, \ldots, S_N$) is a D-dimensional vector. Here, D is the optimization parameter's count. Once the initialization process is done, the population of the positions (solutions) is respect to the iterated cycles, $C = 1, 2, \ldots, MCN$, of the search processes of the employed bees, the onlooker bees and the scout bees. An employed bee generates changes on the position (solution) in her memory according to the local information (visual information) and verifies the nectar amount (fitness value) of the new source (new solution). If in case, new source's quantity is greater than the earlier one, the bee memorizes the new position and leaves the old one, or else, it will maintain the earlier one in her memory. The nectar information's food source and their location will be shared with the onlooker bees, once the employed bees finish their searching process. From the received information, onlooker bee computes the nectar information and selects a food source with a probability based on its nectar amount.

According to employed bee, it will change the position in its memory and verifies the nectar amount of the candidate source. If the nectar is greater than the earlier one, it will memorize the new one and leaves the old one. The primary step of the algorithm is:

1: Initialize Population

2: repeat

3: Place the employed bees on their food sources

4: Place the onlooker bees on the food sources depending on their nectar amounts

5: Send the scouts to the search area for discovering new food sources

6: Memorize the best food source found so far

7: until requirements are met

In this algorithm, every cycle has three steps: compute the nectar amount by sending the employed bees onto their food sources, then shares the nectar information with onlooker bee and it will choose the food source regions based on its nectar amount of the food sources; defining the scout bees and then transmitting without any order onto the possible new food sources. In the stage of initialization, bees will choose the group of food sources is randomly and nectar amount is defined. In 1st cycle, these bees arrives hive and exchange the nectar information of the sources with bees that were waiting in the dancing area. Usually the bees will wait in the dance area mainly to take a decision to select the food source is termed as onlooker and the employed bee will visit the food source by herself. Once after sharing the details with onlooker bee, every employed bee will visit the food source at the previous cycle because position of the food source is there in her memory and then selects a new food source through the visual information in the neighbourhood, which is there in her memory and she will compute the nectar amount for new food source. 2nd step, an onlooker chooses a food source area which is based on the nectar information spread by the employed bees on the dancing area. Probability of the selecting the food source raises when the nectar amount of a food source gets maximized. Deciding the new food source is proceeded by the bees by distinguishing the position food source process visually. At 3rd step, if the nectar amount of the food source is left alone, a new one is defined in any order by the scout bee and it will be altered by the left one. In our approach, at every cycle, just one scout bee go out in searching of food and the employed and onlooker bees count should be equal to everyone. These three steps are iterated through a fixed number of cycles which is known as Maximum Cycle Number (MCN) or just before fulfilling the termination strategy. An artificial onlooker bee selects a food source based on the likelihood value correlated with food source, pi, is computed through,

$$p_i = \frac{fit_i}{\sum_{n=1}^{SN} fit_n} \quad (6)$$

where fit_i is the fitness value of the solution i which is proportional to the nectar amount of the food source in the position I and SN denotes the amount of food sources which is equal to the employed bees or onlooker bees count. To generate a candidate food position from the old one in memory, the ABC utilizes:

$$v_{ij} = x_{ij} + \phi_{ij}(x_{ij} - x_{kj}) \quad (7)$$

where k ∈ {1, 2, . . . ,SN} and j ∈ {1, 2, . . . ,D} are indexes, which is selected without any order, it has to be vary from i. ϕ_{ij} is a random number between [-1, 1]. It manages the neighbour food sources production, around $x_{i,j}$ and it indicates the position of food source which is distinguished visually by a bee. Equation 7, indicates the variations among the parameters of the $x_{i,j}$ and $x_{k,}$ minimizes, the perturbation on the position xi;j gets minimized. So, the search method access the optimum solution in the search space, the step length is minimized flexibly. If a parameter value produced by this task goes beyond its predetermined boundary that value will be fixed to an acceptable value. In our work, the parameter value goes beyond the boundary to fix its boundary limits. The food source of which the nectar is left alone by the bees is iterated with a new food source by the scouts. In ABC, this is simulated by generating the position without any order and altering it with the left one. In ABC, if a position can't be enhanced by a predetermined cycle, then it is considered to be deserted. In ABC algorithm, predetermined value cycle is a significant control parameter, which is termed as ''limit" for forsaking. Let's consider the deserted source as xi and j ∈ {1, 2, . . . ,D}, then the scout identifies a new food source to be altered with x_i . This operation can be determined as

$$x_i^j = x_{min}^j + \mathrm{rand}[0,1](x_{max}^j - x_{min}^j) \quad (8)$$

After generating the candidate source position of the vi;j , it will be computed by artificial bee, its achievement is distinguished with its old one. If the new food source nectar amount is equal or better than the old source, it will be altered with the old one in the memory. Else, old one will be maintained. Specifically, a greedy selection mechanism is enforced as the selection task among the old and the candidate one. Absolutely, ABC algorithm applies four different selection processes: (1) a global probabilistic selection process, in which the probability value is computed by equation 6 and it is utilized by the onlooker bees for identifying hopeful areas, (2) a local probabilistic selection process proceeded in the area through the employed bees and the onlookers based on the visual information namely colour, shape and fragrance of the flowers (sources) (bees won't be able to detect the nectar source variety till they reach at the proper location and segregate between sources rise according to the scent) for defining a food source about the source in the memory which is explained in equation 7. (3) a local selection termed as greedy selection process proceeded by onlooker and employed bees in that if the nectar amount of the candidate source is comparatively good than that of the current one, the bee leave the current one and memorizes the candidate source created by equation 7. Else, the bee maintains the current one in its memory. 4) A random selection process proceeded by scouts as determined in (8).

It is understood from the aforementioned description, that there are control parameters are basic in AMC algorithm: Food source count is equal to the employed or onlooker bees (SN) count, the boundary value and the maximum cycle number (MCN). In honeybees, the hiring rate indicates the measure of how soon the bee colony identifies and use a fresh food source. Likewise artificial hiring indicates the measurement quickly identify the appropriate solution and good quality solutions of the complex optimization issue. The durability and growth of the bee colony depends on the accelerated discovery and effective usage of the best food resources. Likewise, the successful solution of complex engineering issue is linked to the almost fast while identify the better solution in a practical manner for the issue which is required to be resolved in the real time. In a robust search process, exploration and exploitation process should be proceeded together. In the ABC algorithm, onlookers and employed bees proceed the exploitation in the search space, the scouts manages exploration process. Elaborated pseudo-code of the ABC algorithm is given below:

1: Initialize the population of solutions x_i, i = 1, . . . ,SN

2: Evaluate the population

3: cycle = 1

4: repeat

5: Produce new solutions t_i for the employed bees by using (7) and evaluate them

6: Apply the greedy selection process for the employed bees

7: Calculate the probability values P_i for the solutions x_i by (6)

8: Produce the new solutions t_i for the onlookers from the solutions x_i selected depending on P_i and evaluate them

9: Apply the greedy selection process for the onlookers

10: Determine the abandoned solution for the scout, if exists, and replace it with a new randomly produced solution x_i by (8)

11: Memorize the best solution achieved so far

12: cycle = cycle + 1

13: until cycle = MCN

Defining the Fuzzy Membership Functions

Deciding the membership functions is the main step before rectifying the quadratic programming of RMC-FSVM. It demands good discriminatory power, while selecting membership function; it also demands the domain expert's knowledge. In our work, assume three strategies for deciding the membership function, which is represented as SVM, linear, and logistic. First define the primary score for every data record by a basic scoring method which is represented as score k for record k. For the SVM strategy, the standard SVM is utilized to produce the primary score according to the distance among the data point and hyper plane, while linear and logistic regressions were utilized for the linear and logistic approaches, correspondingly. Utilizing the upcoming linear map function to map the primary scores into the membership degree in the unit interval [0, 1]:

$$\text{Membership } S_k = \frac{\text{score}_k - \min\limits_{k=1,..N} \text{score}_k}{\max\limits_{k=1,..N} \text{score}_k - \min\limits_{k=1,..N} \text{score}_k} \quad (10)$$

where $\min_{k=1,..N} \text{score}_k$ and $\max_{k=1,..N} \text{score}_k$ are the minimum and maximum of the scores of the dataset, correspondingly.

4. Experimentation Results

The execution of the proposed HMRRMC-FSVM-ABC method is computed in this section and also distinguished with that of the exact MRRMC-FSVM, RMC-FSVM, FSVM and SVM methods. The experiments were proceeded on two real-world datasets. They are: Wisconsin Diagnostic Breast Cancer (WDBC) dataset and PD600, a group of loan application data from Taiwan. To create a fair comparison, entire support vector classifiers integrated a Gaussian kernel, and the same MATLAB quadratic programming solver is chosen for entire methods. The codes are implemented in MATLAB R2011a on a 3.40-GHz Intel Core i7–3770 CPU with 16-GB RAM running Windows Server 2008.

Datasets

Two datasets were used for implementing this methodology. WDBC and PD600. For the WDBC dataset, there are 683 instances after eliminating those with missing values. Each and every instance has nine attributes: clump thickness, uniformity of cell size, uniformity of cell shape, marginal adhesion, single epithelial cell size, bare nuclei, bland chromatin, normal nucleoli, and mitoses. All these attributes are numeric and normalized to a value among 0 and 1. The classification work is estimated whether a tumor is benign (1) or malignant (2). The monotonic attributes were defined by approaching a senior medical doctor.

The PD600 dataset has 600 loan applications given by local bank in Taiwan, covering the period from 2001 to 2002. Every application has 17 variables providing the information of a applicant’s sex, age, marital status, work location, years in the present job, education, type of company, real estate, monthly salary ($NT), family has an idea about the loan application, real income, monthly balance ($NT), record of payment, number of times record checked, utilization of revolving interest, credit ($NT), and operations. After questioning the bank management, the following variables were chosen according to a review of the literature and materials related to Joint Credit

Information Center Certification: the data on a personal credit application form, the borrower's bankbook, wage transfer accounts, household registration certification, and tax withholding voucher. The work is to select the loan applicants who will satisfy their loan responsibilities (non-default: 2) and those who fail to do so (default: 1).

Performance Measures

The proposed HMRRMC-FSVM-ABC algorithm is distinguished with the other support vector classifiers, analyzing their execution with respect to the accuracy, recall, precision, F-measure.

Accuracy is the most intuitive measurement criterion, which exactly explains the capability of prediction according to the proportion of the verified data that are exactly classified, and this is determined as follows:

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{FP} + \text{TN} + \text{FN}} \quad (11)$$

Where TP (true positives), FP (false positives), TN (true negatives), and FN (false negatives) are the normal values utilized to define the predictive power of a method in data mining, and these are determined like:

True positives (TP) = number of instances with positive outcomes that are correctly classified.

False positives (FP) = number of instances with positive outcomes that are wrongly classified.

True negative (TN) = number of instances with negative outcomes that are correctly classified.

False negative (FN) = number of instances with negative outcomes that are wrongly classified.

Recall measures the proportion of actual positives that are exactly detected and this is determined as follows:

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad (12)$$

The precision rate is the proportion of test instances with positive predictive result which are exactly estimated. It is a significant measure of a predictive method, since it give back the likelihood that a positive test give back the basic condition which has be verified. It is determined as

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \quad (13)$$

The conventional F-measure is the harmonic mean of precision [positive predictive value (PPV)] and recall (sensitivity) and is determined as

$$\text{F} - \text{measure} = 2 \cdot \frac{\text{PPV} \cdot \text{Sensitivity}}{\text{PPV} + \text{Sensitivity}} \quad (14)$$

According to the attention of the F-measure, both precision and recall were considered to remove the situation with greater precision and low recall, or vice versa. In this manner, the proposed methodology can be distinguished performance of various algorithms.

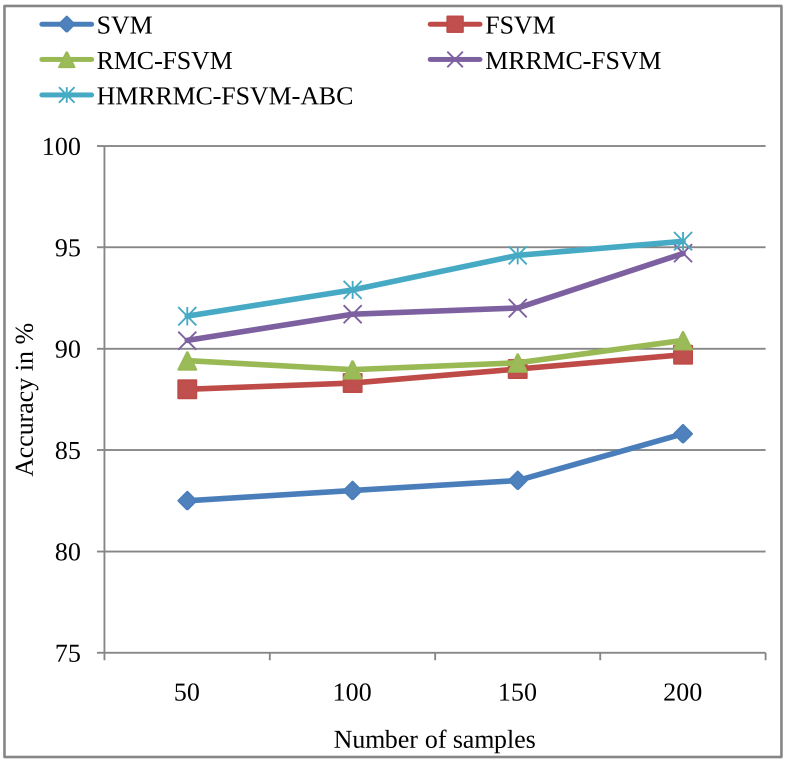

Figure 1: Accuracy Comparison of Classifiers

Fig 1.indicates that the pictorial representation of accuracy when distinguished with the various classifiers. It provides that the proposed HMRRMC-FSVM-ABC and distinguished with current MRRMC-FSVM, RMC-FSVM, FSVM and SVM. The proposed system has accomplished greater accuracy than the current classifiers for two real datasets, since it reduces the error correlated with virtual sample by utilizing the prior knowledge of monotonicity and the learning process has capable of accomplishing better monotonicity in order to rise the accuracy.

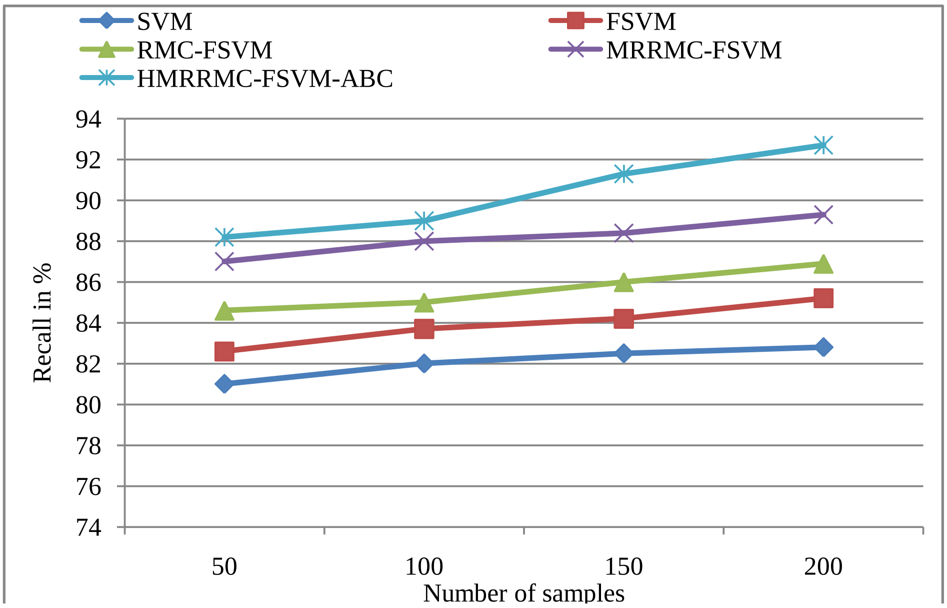

Figure 2: Recall Comparison of Classifiers

Fig2. Provides the pictorial representation of recall distinguished among various classifiers. It provides the proposed HMRRMC-FSVM-ABC when compared with current MRRMC-FSVM, RMC-FSVM, FSVM and SVM. The proposed system accomplishes greater recall than the current classifiers for two real datasets.

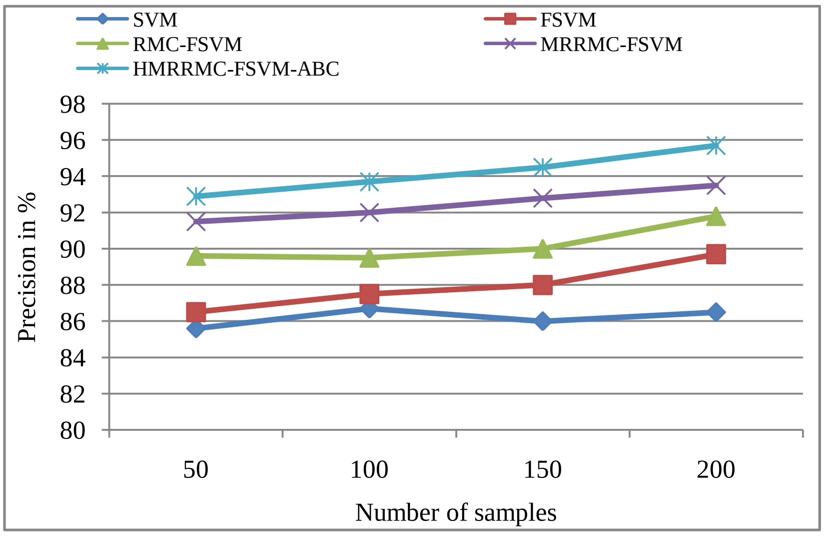

Figure 3: Precision Comparison of Classifiers

Fig 3.indicates that the pictorial representation of precision when compared with various classifiers. It provides the proposed MRRMC-FSVM and when compared with presents RMC-FSVM, FSVM and SVM. The proposed system accomplishes greater precision than the present classifiers for two real datasets.

5. Conclusion

Prior knowledge incorporation and learning is the complex work, which require effective learning knowledge appropriately, also learning process should be implemented with more care, since it is a complex task. By concentrating on RMCFSVM, proposed method rectifies these issues; proposed method has relation with building monotonicity constraints by bringing-in the optimal framework such as Hybrid MRRMC-FSVM with Artificial Bee Colony (ABC) (HMRRMC-FSVM-ABC). Here, IT will automatically choose the select virtual pairs which were defined according to the minimum and maximum of certain random subsets of the input dataset. This selection process is done by artificial bee colony (ABC) procedure with minimum and maximum of specific random subsets of the input dataset, which is assumed as the fitness value of ABC.. This will give back the entire monotonicity and provide good output for the experiment which is proceed to show any important variations for higher number of monotonicity constraints. The computation of the proposed research method is implemented in matlab simulation environment, which confirms that the proposed methodology heads and provides better result when compared to the current methods with respect to the accurate knowledge learning.

References

[1] Han J., Pei J., Kamber M., Data mining:concepts and techniques, Elsevier (2011).

[2] Han J., Cheng H., Xin D., Yan X., Frequent pattern mining: current status and future directions, Data Mining and Knowledge Discovery 15(1) (2007), 55-86.

[3] Boran F.E., Genç S., Kurt M., Akay D., A multi-criteria Intuitionistic fuzzy group decision making for supplier selection with TOPSIS method, Expert Systems with Applications 36(8) (2009), 11363-11368.

[4] Domingos P., Toward knowledge-rich data mining, Data Mining and Knowledge Discovery 15(1) (2007), 21-28.

[5] Afantenos S., Karkaletsis V., Stamatopoulos P., Summarization from medical documents: a survey, Artificial intelligence in medicine 33(2) (2005), 157-177.

[6] Alcalá-Fdez J., Fernández A., Luengo J., Derrac J., García S., Sánchez L., Herrera F., Keel data-mining software tool: data set repository, integration of algorithms and experimental analysis framework, Journal of Multiple-Valued Logic & Soft Computing 17 (2011).

[7] Yang Q., Wu X., 10 challenging problems in data mining research, International Journal of Information Technology & Decision Making 5(04) (2006), 597-604.

[8] Markus M.L., Majchrzak A., Gasser L., A design theory for systems that support emergent knowledge processes, MIS quarterly (2002), 179-212.

[9] Aggarwal C.C., Data classification: algorithms and applications, CRC Press, Edition (2014).

[10] Li S.T., Chen C.C., A regularized monotonic fuzzy support vector machine model for data mining with prior knowledge, IEEE Transactions on Fuzzy Systems 23(5) (2015), 1713-1727.

[11] Miles J.C., Moore C.J., Practical knowledge-based systems in conceptual design, Springer Science & Business Media (2012).

[12] Duivesteijn W., Feelders A., Nearest neighbour classification with monotonicity constraints, Machine Learning and Knowledge Discovery in Databases (2008), 301-316.

[13] Falck T., Suykens J.A., De Moor B., Robustness analysis for least squares kernel based regression: an optimization approach, Proceedings of the 48th IEEE Conference on 28th Chinese Control Conference Decision and Control (2009), 6774-6779.

[14] Evgeniou T., Boussios C., Zacharia G., Generalized robust conjoint estimation, Marketing Science 24(3) (2005), 415-429.

[15] Doumpos M., Zopounidis C., Monotonic support vector machines for credit risk rating, New Mathematics and Natural Computation 5(03) (2009), 557-570.

[16] Pelckmans K., Suykens J.A., Van Gestel T., De Brabanter J., Lukas L., Hamers B., Vandewalle J., LS-SVM lab: a mat lab/c toolbox for least squares support vector machines, Tutorial, KU Leuven-ESAT, Leuven, Belgium 142 (2002), 1-2.

[17] Potharst, R., Feelders, A.J.,Classification trees for problems with monotonicity constraints, ACM SIGKDD Explorations Newsletter 4(1) (2002), 1-10.

[18] Greco S., Matarazzo B., SłowiĚski R., In: Zopounidis, C. (ed.) Operational Tools in the Management of Financial Risks, Kluwer Academic Publishers, Dordrech (1998), 121-136.

[19] Gamarnik D., Efficient learning of monotone concepts via quadratic optimization, Proc. of the eleventh annual conference on computational learning theory, ACM Press, New York (1998), 134–143.

[20] Pendharkar P.C., Rodger J.A., Technical efficiency-based selection of learning cases to improve forecasting accuracy of neural networks under monotonicity assumption, Decision Support Systems 36(1) (2003), 117-136.

[21] Doumpos M., Pasiouras F., Developing and testing models for replicating credit ratings: A multicriteria approach, Computational Economics 25(4) (2005), 327-341.

[22] Pazzani M.J., Mani S., Shankle W.R., Acceptance of rules generated by machine learning among medical experts, Methods of information in medicine 40(5) (2001), 380-385.

[23] Daniëls H.A.M., Kamp B., Application of MLP networks to house pricing and bond rating, Neural Computing and Applications 8(3) (1999), 226-234.

[24] Wang S., Adaptive non-parametric efficiency frontier analysis: a neural-network-based model, Computers & Operations Research 30(2) (2003), 279-295.

[25] Yibing T.J.S., New method of analog circuit fault diagnosis using fuzzy support vector machine,Journal of Electronic Measurement and Instrument 6 (4) (2009).

[26] Ekanayake J., Li H., Zhang B., Gunarathne T., Bae S.H., Qiu J., Fox G., Twister: a runtime for iterative map reduce, Proceedings of the 19th ACM international symposium on high performance distributed computing (2010), 810-818.

[27] Karaboga D., Basturk B., A powerful and efficient algorithm for numerical function optimization: artificial bee colony (ABC) algorithm, Journal of global optimization 39(3) (2007), 459-471.