

A Survey of Modified Support Vector Machine using Particle of Swarm Optimization for Data Classification

By

Rahul Malviya, Asstt Prof.sushil Tiwari, Prof.S.R.Yadav

Errahulmalviya@yahoo.in, sushiltiwari24@gmail.com, Techmillenniumk.yadav@gmail.com

ABSTRACT

Dynamic feature evaluation and concept evaluation is major challenging task in the field of data classification. The continuity of data induced a new feature during classification process, but the classification process is predefined task for assigning data into class. Data comes into multiple feature sub-set format into infinite length. The infinite length not decided the how many class are assigned. Support vector machine is well recognized method for data classification. For the process of support vector machine evaluation of new feature during classification is major problem. The problem of feature evaluation decreases the performance of Support Vector Machine (SVM). For the improvement of support vector machine, particle of swarm optimization technique is used. Particle of swarm optimization controls the dynamic feature evaluation process and decreases the possibility of confusion in selection of class and increase the classification ratio of support vector machine. Particle of swarm optimization work in two phases one used as dynamic population selection and another are used for optimization process of evolved new feature.

Keywords: Feature Selection, Support Vector Machine, particle of Swarm Optimization.

I. INTRODUCTION

Data classification is more challenging task because of unique properties of data stream. some unique properties are Data streams have infinite length ,due to this impractical to store and use historical data for transferring .that's why traditional multipass learning algorithm are not applicable to data streams. next one is the concept drift observed in data stream,it occur when concept of data changes over time.The classification model continuously adapt itself to recent concept drift,to address the concept drift and the other one is the data stream also uses the concept of evolution which only occur when a novel classes appear in the data stream.with advanced data stream technologies[1]. Collecting large amounts of data in various application domain example Stock market fluctuations ,transaction of credit card, web click stream, network traffic monitoring, position updates of moving objects in location based services and text streams from news[2]. Data stream mining studied intensively in past few years because of its potential in industry application. In general first learn one or multiple classification models from the past records of evolving data and after that selected a model that matches to the current data to predict the new data records .All the existing techniques present in the data stream classification assumed that at each time stamp there are both type of training data for learning is available which is called as positive training data or negative training data. the aim of data stream classification is learn from the past data and classification. First one is the data stream have infinite length .Second, data stream observed concept-drift.Third, the concept of evolution. Classification in data stream is of two type it may be single model incremental approaches or Ensemble technique .In Ensemble technique the classification output is a

function of prediction of different classifiers. This technique is more popular than single model counterparts because of simple implementation and higher efficiency. Ensemble technique uses a chunk based approach for learning. In this divide the data into chunks and uses or trained a model from one chunk. To classify the unlabeled data use the chunk based model. Chunk based keep the fixed sized ensemble, in which older model is replaced by the new model or updated model. Some chunk based technique can't detect recurrent classes also.

2. SYSTEM MODEL

Support Vector Machine (SVM) is a Supervised learning approach in which studied the data and verify its pattern, used for classification analysis and regression analysis assume a set of training data points each label as employing to one of more with in more groups. Supervised learning involve learning from a training set of data. it is machine learning task of inferring a function from labeled training data. In supervised learning, each example is a pair consisting of an input object and a desired output value. it analyzes the training data and produces an inferred function which is used for mapping new example. A Support Vector Machine training data set method construct a model to apply new example as points in space, mapped so that the example of the separated categories are divided by a clear gap that is as wide as possible. new example are mapped into the same space and predicted to belong to a category based on which side of the gap they fall on. some of the application of SVM are Use in text and hyper text categorization. Used in classification of images. Useful in medical science to classify protein. Hand written character can be recognized by using SVM

Stastical learning theory is a frame work for machine learning in the fields of statics and function analysis. Stastical learning deal with problem of finding a predictive function based on data. it is also a tool for creating practical algorithm for pattern reorganization. its application are computer vision, speech reorganization, bioinformatics. with help of theoretical analysis discover a general model of generalization. On the basic of the VC dimension concept, constructive distribution –independent bounds on the rate of convergence of learning process can be obtained and the structural risk minimization principle has found. The new understanding also changes the algorithm approaches to pattern reorganization.

Particle swarm optimization (PSO) is a population based stochastic optimization technique developed by Dr. Eberhart & Dr. Kennedy in 1995, inspired by social behavior of bird flocking or fish schooling. it is a computational method that optimizes a problem by iteratively to find the best possible solution. The power of this technique is computation and sharing of information within the algorithm. its internal communication is based from the social behavior of individuals. Individuals are termed as particle henceforth are flown through the multi-dimensional search space and all the particle represent the possible solution to the problem. solution fitness is depend on the multi objective performance function related to optimization problem being solved. Particle movement is influenced by two factor which are using information from iteration-to-iteration information, particle stores best solution in its memory called p best and experiences an attraction towards solution as it traverses through the solution search space. as a result, particle stores best solution in its memory visited by any particle & experiences an attraction towards this solution called the g best. The first and second factor are called cognitive and social components respectively. at the time of iteration, the p best and g best are updated if better solution is found which is in term of fitness. this process is continuously performed until the desired result is found.

3. PREVIOUS WORK

Supervised classification has been used in this research to categorize the data as the SIMPLiCity dataset contains images with well-defined labels. Input dataset to the supervised classifier are label as Africa, beach, buildings, buses, dinosaurs, elephants, flowers, horses, mountains and food etc [1]. The feature selections of the data, in the WH format are given to the supervised classifier to infer a rule that assigns a label to each label. The classifier assigns a class label to the output value y , that best matches the given input pattern and is denoted by C_k , $k = 1, 2, \dots, K$, where K is the number of classes. SVM (Support Vector Machine) are a set of expected supervised machine learning methods used for classification, retrieval and regression as well as image

processing [6]. They belong to a family of generalized linear and non-linear classification and regression. A special characteristic of SVM (Support Vector Machine) is that simultaneously minimizes the predictive classification and regression error and maximizes the geometric hyper plane margin in data classification [4]. Hence SVM (Support Vector Machine) is known as Maximum Margin Classifiers (MMC). Generally, SVM is totally based on the Structural Risk Minimization (SRM) in data classification [4]. SVM maps the different input vector from data set to a higher dimensional space in image processing, where a maximum dividing hyper plane is built for desired dataset classification [7]. We are built two different parallel hyper planes on each edge of the hyper plane that divide the dataset nodes in dataset [9]. In Dataset classification, the dividing hyper plane is the desired hyper plane which maximal the distance between the two different parallel hyper planes. An predictive is made which the maximal the margin or distance among these parallel hyper planes being more best the generalization error of the linear classifier supervised machine learning [8]. Then SVM finds a linear as well as non-linear separating hyper plane with the maximal margin in this higher dimension space of data classification. Basically, $C > 0$ is the penalty parameter of the generalization error term in data classification. Consequently, $\text{Ker}(x_i, x_j) \equiv \Phi(x_i)^T \Phi(x_j)$ is called the main kernel function[10]. There are many kernel mappings in Support Vector Machine supervised learning, so how to choose a good kernel mapping is also a research topic in current chapter. Hence, for basic purposes, there are some different popular kernel mappings:

- Linear kernel: The linear kernel equations is $\text{Ker}(x_i, x_j) = x_i^T x_j$.
- Polynomial kernel: The poly kernel is $\text{Ker}(x_i, x_j) = (\gamma x_i^T x_j + r)^d$, $\gamma > 0$
- RBF kernel: The RBF kernel is $\text{Ker}(x_i, x_j) = \exp(-\gamma \|x_i - x_j\|^2)$, $\gamma > 0$

- Sigmoid kernel: The sigmoid equations is $\text{Ker}(x_i, x_j) = \tanh(\gamma x_i^T x_j + r)$

Multiple classifiers are fused at the decision level, and the particle swarm optimization algorithm finds optimal decision threshold for each classifier and the optimal fusion rule[11]. Specifically, we present an optimal fusion strategy for fusing multiple classifiers to satisfy accuracy performance requirements, as applied to a real-world classification problem. The optimal decision fusion technique is found to perform significantly better than the conventional classifier fusion methods, i.e., traditional decision level fusion and averaged sum rule[11].

The optimized decision level fusion is found to be performing better than the traditional decision level fusion. Often decision level fusion techniques are considered as suboptimal when compared to other fusion approaches. Such suboptimal performance is due to the two-layered optimization approach adopted by researchers while using decision fusion. Also results presented show a promise of decision level fusion as compared to averaged sum rule. Also it is interesting to find that the combination of highly contrasting fusion rules as “AND” and “OR” result in better performance than the averaged sum rule. Particle swarm optimization problem is used to achieve the optimal decision fusion strategy, which is an intractable problem[9].

4. PROPOSED METHODOLOGY

Feature optimization and data classification in stream data mining is challenging task for researchers for controlling new feature evolution of incoming data for classification. The new evolution of feature compromised the process of proper class mapping. The mapping of class is not perform the performance of classifier is degraded. For the improving the classification rate of stream data in current decade proposed a support vector machine method for evaluation of new feature and improvement of classification[6]. For maintaining a new evolve feature

attribute of data for classification one problem automatically recall such problem is called data drift. Data drift induced a discontinuity of stream data and suffered a problem of classification. For a lack of data drift and feature evolving concept technique support vector machine degrade the performance of classifier in mode of single class and multi-class. For fulfillment of researcher gap in stream data classification proposed an optimized feature selection method for stream data classification. The optimized feature selection method basically based on particle of swarm optimization. Particle of swarm optimization algorithm well knows population based multi-stage optimization technique[11]. The algorithm for proposed method MSVM-PSO is Explained in two phase, which is as follows:

Phase I: Algorithm of PSO (For select optimal feature from data set)

Step 1: All particles' velocity and position are randomly place to within pre-defined ranges.

Step 2: Velocity update – At every iteration, the velocities of all particles are updated based on below expression

$$v_i = v_i + c_1 R_1 (p_{i,best} - p_i) + c_2 R_2 (g_{i,best} - p_i) \dots(1)$$

where p_i is the position and v_i are the velocity of particle i , $p_{i,best}$ and $g_{i,best}$ is the position with the 'best' objective value found so far by particle i and the entire population respectively; w is a parameter controlling the dynamics of flying; R_1 and R_2 are random variables in the range [0,1]; c_1 and c_2 are factors controlling the related weighting of equivalent terms. The random variables facilitate the PSO with the ability of stochastic searching.

Step 3: Position updating – The positions of all particles are updated according to,

$$p_i = p_i + v_i \dots(2)$$

Following updating, p_i should be verified and limited to the allowed range.

Step 4: Memory updating – Update $p_{i,best}$ and $g_{i,best}$ when condition is met,

$$p_{i,best} = p_i \quad \text{if } f(p_i) > f(p_{i,best})$$

$$g_{i,best} = g_i \quad \text{if } f(g_i) > f(g_{i,best}) \dots(3)$$

Where $f(x)$ is to be optimized and it is a objective function.

Step 5: Stopping Condition – The algorithm repeats steps 2 to 4 until certain stopping circumstances are met, such as a pre-defined number of iterations. Once closed, the algorithm reports the values of g_{best} and $f(g_{best})$ as its solution[8].

PSO utilizes several searching points and the searching points gradually get close to the global optimal point using its pbest and gbest. Primary positions of pbest and gbest are dissimilar However, using thee different direction of pbest and gbest, all agents progressively get close up to the global optimum.

Phase II: Algorithm for MSVM (For classification of input data set)

Step 6: candidateSV = { nearest pair from different labels }

Step 7: **while** there are violating nodes **do**

Find a node violator

Candidate_SV = candidate_SV \setminus S violator

Step 8: **if** any $ap < 0$, addition of c to S **then**

candidate_SV = candidate_SV \setminus p

continue till all nodes are pruned

Step 9: **end if**

end while

5. IMPLEMENTATION

In this section discuss experimental process of proposed MSVM-PSO technique for modified support vector machine algorithm for data classification. The proposed method implements in MATLAB 7.8.0 and tested with very reputed data set from UCI machine learning research center. In the research work, I have measured F new, M new and Error rate of classification. To evaluate these performance parameters I have used three datasets from UCI machine learning repository [18] namely crop dataset, Glass dataset and forest fire dataset. Out of these three dataset, one is small dataset namely glass dataset; and remaining two are large datasets namely crop dataset and forest fire dataset.

6. CONCLUSIONS

We have presented a novel approach is MSVM-PSO technique. The proposed technique finds the classified data from large input data chunk, which is simple and effective. More importantly, the PSO optimization method is used in a desired way to make full use of optimize the data set between neighborhood features for optimization trends. Experiments explained that the proposed technique can well keep the original and complementary information of all input data set. Encouragingly, the proposed technique is very robust to

data classification. Furthermore, the proposed approach is computationally effective, making it quite qualify for desired applications. At last, how to get improve the effectiveness of the concern proposed method by adaptively choosing the parameters of the classification technique can be further researched.

6. FUTURE WORK

The proposed method modified support vector machine solved the problem of feature classification and concept evaluation. The controlled feature evaluation classification process increases the value of F new and M new and reduces the error rate. The particle of swarm prototype cluster faced a problem of right number of cluster, in future used self optimal clustering technique along with particle of swarm optimization.

The future work is as follows

1. We can use other classification technique like Baye's theorem, K-Nearest neighbor, Decision tree and ANN
2. We can use other optimization technique like ACO, COACO, MMACO, Genetic Algorithm and GACO method

7. REFERENCES

- [1] Mohammad M. Masud, Qing Chen, Latifur Khan, Charu C. Aggarwal, Jing Gao, Jiawei Han, Ashok Srivastava "Classification and Adaptive Novel Class Detection of Feature-Evolving Data Streams" TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING, IEEE 2012. Pp 1-14.
- [2] Chang-Dong Wang, Jian Huang La, Dong Huang, Dong Huang "SVStream: A Support Vector-Based Algorithm for Clustering Data Streams" IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING, VOL. 25, 2013. Pp 1410-1425.

- [3] Xiao-Li Li, Philip S. Yu, Bing Liu, See Kiong Ng “Positive Unlabeled Learning for Data Stream Classification” SIAM, 2010. Pp 259-270.
- [4] Mohammad M. Masud, Qing Chen, Jing Gao, Latifur Khan, Jiawei Han, Bhavani Thuraisingham “Classification and Novel Class Detection of Data Streams in a Dynamic Feature Space” ECML PKDD, 2010. Pp 337-352.
- [5] Kalyan Veeramachaneni, Weizhong Yan, Kai Goebel, Lisa Osadciw “Improving Classifier Fusion Using Particle Swarm optimization” Proceedings of the IEEE Symposium on Computational Intelligence in Multi criteria Decision Making, 2007. Pp 128-136.
- [6] Tahseen Al-Khateeb, Mohammad M. Masud, Latifur Khan, Charu Aggarwal, Jiawei Han, Bhavani Thuraisingham “Stream Classification with Recurring and Novel Class Detection using Class-Based Ensemble” 2012. Pp 1-10.
- [7] Albert Bifet, Geoff Holmes, Bernhard Pfahringer, Richard Kirkby, Ricard Gavalda “New Ensemble Methods For Evolving Data Streams” 2010. Pp 1-9.
- [8] Peng Zhang, Xingquan, Zhu LiGuo “Mining Data Streams with Labeled and Unlabeled Training Examples” Ninth IEEE International Conference on Data Mining, 2009. Pp 627-636.
- [9] Ashfaqur Rahman , Brijesh Verma “Novel Layered Clustering-Based Approach for Generating Ensemble of Classifiers” IEEE TRANSACTIONS ON NEURAL NETWORKS, VOL. 22, 2011. Pp 781-792.
- [10] Xin Xu, Wei Wang, Guilin Zhang, Yongsheng Yu “An Adaptive Feature Selection Method for Multi-class Classification” 2010. Pp 225-230.
- [11] Kalyan Veeramachaneni, Weizhong Yan, Kai Goebel, Lisa Osadciw “Improving Classifier Fusion Using Particle Swarm optimization” Proceedings of the IEEE Symposium on Computational Intelligence in Multi criteria Decision Making, 2007. Pp 128-136.
- [12] Mohammad M. Masud, Qing Chen, Jing Gao, Latifur Khan, Jiawei Han, and Bhavani Thuraisingham “Classification and Novel Class Detection of Data Streams in a Dynamic Feature Space” in ISMIS 2009, LNAI 5722, Pp 552-558.
- [13] X. Li, P. S. Yu, B. Liu, and S.-K. Ng. Positive unlabeled learning for data stream classification. In Proc. SDM, Pp 257–268, 2009.
- [14] M. M. Masud, Q. Chen, J. Gao, L. Khan, J. Han, and B. M. Thuraisingham. Classification and novel class detection of data streams in a dynamic feature space. In Proc. ECML PKDD, volume II, Pp 337–352, 2010.
- [15] P. Zhang, X. Zhu, J. Tan, and L. Guo, “Classifier and Cluster Ensembles for Mining Concept Drifting Data Streams,” Proc. 10th Int’l Conf. Data Mining, 2010.
- [16] X. Zhu, P. Zhang, X. Lin, and Y. Shi, “Active Learning from Stream Data Using Optimal Weight Classifier Ensemble,” IEEE Trans. Systems, Man, Cybernetics Part B, vol. 40, no. 6, Pp 1607- 1621, Dec. 2010.