



Classifying Limited Resource Data Using Semi-supervised SVM

Naralasetti Veeranjanyulu¹, Jyostna Devi Bodapati^{2*}, Suvarna Buradagunta²

¹ Department of IT, Vignan Foundation for Science Technology and Research, Vadlamudi 522213, Andhra Pradesh, India

² Department of CSE, Vignan Foundation for Science Technology and Research, Vadlamudi 522213, Andhra Pradesh, India

Corresponding Author Email: bjd_cse@vignan.ac.in

<https://doi.org/10.18280/isi.250315>

Received: 27 January 2020

Accepted: 29 March 2020

Keywords:

supervised learning, Laplacian SVM, semi-supervised learning, SVM-light, S³VM

ABSTRACT

The success of machine learning models is subjected to the availability of sufficient amounts of training data. Availability of large amounts of labeled data is difficult especially in the domains of image, speech, video etc. Semi-supervised learning is an approach that uses additionally available unlabeled data to improve performance of a model with limited data. To gain better performance with limited training data we suggest semi-supervised SVM models for EEG Signal classification and image classification tasks. We explore multiple approaches to semi-supervised learning based on Support vector machines: Semi-supervised Support vector machines (S^3VM), S^3VM^{light} , SVM-light and label switching based Support Vector Machine (Lap-SVM) for different tasks. Our experiments show that semi-supervised approaches when trained with sufficient unlabeled data can significantly improve performance of the model when compared with its counterpart supervised model. The proposed models are verified on 3 different benchmark data sets. Proposed semi-supervised approach for image classification task show a remarkable 20% improvement over baseline SVM model.

1. INTRODUCTION

Pattern classification has a remarkable importance for achieving impeccable results in the applications like text and web page classification, image and object recognition and speech processing applications [1-3]. The objective of pattern classification is to find the class to which a given sample is belonging to.

Based on the amount of supervised information for training, ML algorithms are classified into semi-supervised, unsupervised, and supervised algorithms. The training data in the case of unsupervised learning contains no supervisory information. Ex: clustering, dimension reduction and density estimation. In supervised learning the data is labeled. Classification and regression come under supervised learning as labels are used during training.

To build a classification model with high accuracy, adequate number of labeled examples (usually huge in number) are required. Acquiring huge labeled data is difficult in certain domains, especially in DNA sequencing, image, and speech processing applications. At the same time plenty of unlabeled data is available in such applications [4]. Aim of this work is to use these abundant unlabeled data to achieve a better classification model.

SVM is the most robust machine algorithm that yields better generalization for the unseen data. Though any other ML algorithm can be adapted to semi-supervised setting, we prefer to use SVM as it is robust with small scale data. Hence in this work we use SVM based algorithms for semi-supervised training. We have taken P% of the training samples with labels and rest of the data is considered as unlabeled. Our experimental results show that the proposed approaches with limited labeled data is better when compared to the supervised

counterparts trained using the same amounts of labeled data. On BCI and G241c data sets SVM light gives better accuracy whereas on COIL2 data set Laplacian SVM outperforms the other methods.

2. RELATED WORK

Semi-supervised learning (SSL) is a learning method that uses sufficiently large unlabeled data together with limited labeled data to design a learning algorithm that performs better than the one trained using labeled data alone. The most used SSL methods are: i) Self or iterative training ii) Graph based techniques iii) SSL for SVM methods and iv) SSL for generative models.

Self-training: Self-training is the earliest and the simple semi-supervised learning procedure [5]. Self-training works on the assumption that predictions of a classifier with high confidence are correct. In this approach, a classifier, usually called as a wrapper method, (this classifier can be any one of the existing classifiers) is trained using limited labeled data. This initial classifier is applied on unlabeled data and then the labeled data is augmented by adding confidently classified unlabeled examples. This process is iterated until some criteria is met.

This is the simplest SSL method, as any existing classifier can be used as a wrapper. Its disadvantage is that the initial mistakes may reinforce themselves [5].

Graph based Methods: These type of SSL methods works based on cluster assumption [6]. In these techniques, data used for training including labeled and unlabeled is represented as vertices in a graph and the weights of edges connecting the vertices represent the level of similarity between the data

points associated with those vertices. Important stages in a Graph-based SSL (GSSL) algorithm are: i) Graph construction and ii) Label inference.

In these graph-based algorithms, type of the graph used, and the similarity measure plays a vital role on the performance of the learning algorithm [7]. Once the graph is available, label inference algorithm is applied on that graph to infer labels for all the unlabeled data points. Semi-supervised learning can be posed as a graph min cut problem [8]. Many popular GSSL algorithms including graph cuts [8-11], graph-based random walks [12, 13], manifold regularization [14, 15], and graph transduction [16, 17] are few of the popular existing graph-based SSL methods.

Advantages: These graph-based SSL methods do not involve building any classification model and majority of these graph-based SSL algorithms involve convex optimization.

Disadvantages: Performance of these algorithms depends on how well the relationship among the data points is represented in the graph. The performance of these algorithms depends on how well the relationships among the data items are captured while constructing the graph [10].

3. PROPOSED SEMI-SUPERVISED SVMs FOR IMAGE CLASSIFICATION

There are many real-time applications like speech recognition, video processing and medical images where there is very limited labeled data is available. With this limited data obtaining good generalization on unseen data is a herculean task. Hence there is a need for introducing models that are more generalizable even when trained using limited labeled data. One possibility is semi-supervised learning. In this work, SVM based semi-supervised learning is used for two different applications: Images and Brain computer interface data.

SVM works on the assumption that separating hyper plane passes through a low density region [18]. SVMs work based on Structural Risk Minimization (SRM) principle by seeking for the maximum separating hyper plane. SVM is more robust [19] compared to the existing ML algorithms as the objective of SVM is to optimize generalization error instead of optimizing the training error, which result in good generalization performance. Let \bar{x}_n be the n^{th} training data point, y_n be the label corresponding to the data point \bar{x}_n , α_n be the Lagrangian coefficient, $k(\cdot, \cdot)$ be the kernel function, NS be the number of support vectors and \bar{x} be any test data point,

$$\bar{w} = \sum_{n=1}^{NS} \alpha_n y_n \phi(x_n) \quad (1)$$

where, $\phi(\tilde{x}_n)$ is the transformed version of \tilde{x}_n in kernel space, the goal is to find a hyper plane [15], $g(\tilde{x}_n)$:

$$\bar{w} = \sum_{n=1}^{NS} \alpha_n k(\bar{x}_n, x_n) = 0 \quad (2)$$

Traditional SVM: The goal of SVM is to seek for a hyper plane with maximum distance from both the classes. The hyper plane produced by SVM is called as the maximum margin hyper plane as it separates the data of either classes with maximum margin. In this approach the loss on data,

$\max(1 - y_n(\bar{w}^T x_n + b), 0)$, is called as hinge loss [18] as the plot of the loss function looks like a hinge.

S³VM: Popularly known as transductive SVM (TSVM) is a semi-supervised SVM that finds a maximum margin hyper plane with respect to training data of both the classes including unlabeled data [20]. Assuming N number of examples used for training, N^L examples are considered to be labeled and the remaining N^U are unlabeled. The boundary obtained by S³VM is the one with minimum generalization error on the unlabeled data [21]. As the labels for unlabeled data are not available, putative labels are identified for the unlabeled data using the standard SVM trained on labeled data and a loss called as hat loss [18] is applied on unlabeled data instead of hinge loss. S³VM optimization problem is:

$$L = \min_{w,b} \left\| \tilde{w} \right\|_2 + C \sum_{n=1}^{N^L} \max(1 - y_n(\tilde{w}^T \bar{x}_n + b), 0) + C^* \sum_{n=N^L+1}^N \max(1 - y_n|\tilde{w}^T \bar{x}_n + b|, 0) \quad (3)$$

The hyper parameters C, C^* in (3) indicate loss on labeled and unlabeled data respectively. Finding the exact solution to the objective function of TSVM is NP-hard and hence approximate solutions are proposed. The early solutions proposed to the S³VM objective problem cannot even handle few hundreds of unlabeled examples [22, 23]. Different approaches like S³VM^{light}, gradient S³VM, Branch and Bound SVM etc. are proposed solutions to the approximated versions of S³VM objective function that are widely used [24]. S³VMs are applicable wherever SVMs are applicable but Unlike SVMs, S³VM objective is non-convex and the exact solution is not available [25].

S³VM^{light}: [25] is a search based algorithm that uses label switching as the base line. Initially a fraction $N^U * r$ of the unlabeled samples is labeled by applying standard SVM. Let y_u denotes the labels for unlabeled examples. While labeling the unlabeled samples a threshold is applied so that only the outputs so that only a fraction r of them are labeled as positive and remaining are labeled as negative. Where r is the prior on positive samples of unlabeled data. In further iterations the labels of either classes are mutually switched so that the constraint on the number of examples would remain satisfied.

Laplacian Support Vector Machines (Lap-SVM): is a straightforward extension of SVM to the SSL domain [14]. The hyper plane produced by lap-SVM is similar to the one derived by standard SVM except that the kernel used is different. Let W is the weight matrix defined as

$$w_{ij} = \frac{e^{-||x_i - x_j||}}{2\sigma^2} \quad (4)$$

D is the diagonal degree matrix defined as

$$d_{ii} = \sum_{i=1}^N w_{ij} \quad (5)$$

p is an integer, $L = D - W$ is the graph Laplacian matrix, L_{norm} is the normalized graph Laplacian matrix, defined as

$$L_{norm} = D^{-1/2} L D^{-1/2} \quad (6)$$

τ_A and τ_I are the ambient and intrinsic norms that act as regularization parameters and their ratio controls the extent of deformation. Now define

$$M = \frac{\tau_I}{\tau_A} L_{norm}^P \quad (7)$$

and define K, a matrix with

$$K_{ij} = k(\tilde{x}_i, \tilde{x}_j) \quad (8)$$

$$\text{and } k_x = (k(\tilde{x}_1, \tilde{x}) \dots k(\tilde{x}_n, \tilde{x}))^T \quad (9)$$

To produce a classification function that is defined for the entire input space, Lap-SVM constructs a kernel \tilde{k} [15, 26] as:

$$\tilde{k}(\tilde{x}, \tilde{z}) = k(\tilde{x}, \tilde{z}) - k_x(I + MK)^{-1}MK_z \quad (10)$$

Different from SVM-light [17, 20, 27], LapSVM is a graph-based approach. It is a non-transductive approach and hence can be applied to unseen data [14]. Due to its simplicity, training LapSVM in the primal can be used to deal with large data sets [28]. Fast Laplacian SVM (FLapSVM) is an extension to Lap-SVM that exhibits better generalization ability [29].

4. EXPERIMENTAL RESULTS

This section reports the performance of semi-supervised SVM and Semi-supervised GMM approaches on different data sets. Each data set is divided into train and test data sets. In each data set, 80% of the data is used for training and the remaining data is used for testing. If L% of the training data is considered as labelled then the remaining data of the training is considered to be unlabeled. In our experiments L value is varied from 10 to 100 with an interval of 20. All the results reported here are on the test data and for the best suitable parameters.

A. Summary of the data sets:

In our studies, the performance of semi-supervised SVM classifier is evaluated on the following three benchmark data sets [5]:

- Brain computer interface (BCI) data set: This data set consists of the EEG signals collected from 400 persons. The data is collected when he was asked to imagine the movements of his hands. Given the EEG signals the task is to identify whether left hand or right hand was used, a binary classification task.
- Columbia object image library (COIL-2): This data set consists of color images 100 different objects. Out of the total samples only 24 objects were randomly sampled and then divided into 2 different classes.
- G241c: It is a synthetic data set that is generated from two isotropic Gaussian's of unit-variance.

Table 1 shows the statistics of the datasets used in experimental studies of the proposed work.

To prove the efficiency of semi-supervised SVMs we compare semi-supervised models with their counterpart non-semi supervised SVMs that uses labeled data alone for training. Semi-supervised models like Self-training, S^3VM^{light} and Lap-

SVM uses unlabeled data along with labeled data to train the model.

In all the experiments, we use a variable 'L' to represent how much portion of the training data is labeled and how much is unlabeled.

Table 1. Summary of the data sets used for s^3vm approaches

Data set	Classes	Features	Samples
BCI	2	117	400
G241c	2	241	1500
COIL2	2	241	1500

Table 2. Performance of semi-supervised SVM models on BCI data set

Data set	L in %	Classification accuracy		
		SVM	S^3VM^{light}	Lap-SVM
BCI	10	50.00	50.00	53.25
	30	48.75	70.00	58.00
	50	66.25	73.75	63.00
	70	62.50	72.50	70.75
	100		67.50	

In Table 2 performance of different semi-supervised algorithms (Self-training, S^3VM^{light} and Lap-SVM) are compared and the results are reported in terms of classification accuracy.

From Table 2 it can be observed that S^3VM^{light} and Lap-SVM performs better than self-training. Performance of S^3VM^{light} is consistent compared to Lap-SVM. On BCI data set with 30% of labeled data S^3VM^{light} reaches the performance of supervised learning with complete training data labeled.

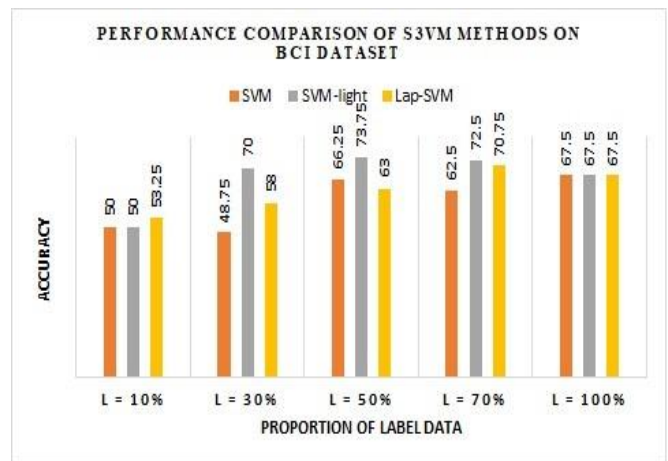


Figure 1. Performance comparison of semi-supervised SVM on BCI data set

Figure 1 shows the results on BCI data set using all the three SSL approaches. We can observe that L=50% gives highest accuracy using SVM-light approach.

Table 3 shows COIL2 data set performance of S^3VM^{light} with 50% of labeled data and Lap-SVM with 30% of labeled data is comparable with the performance of the supervised learning with complete training data labeled. Figure 2 shows the results on G241c data set using all the three SSL approaches. We can observe that L=50% gives highest accuracy using SVM-light approach.

Table 3. Performance of semi-supervised SVMs on G241c data set

Data set	L in %	Classification accuracy		
		SVM	S3VMlight	Lap-SVM
G241c	10	71.33	85.67	50.66
	30	78.33	86.00	62.73
	50	82.66	86.33	68.93
	70	82.66	87.00	71.06
	100		88.67	

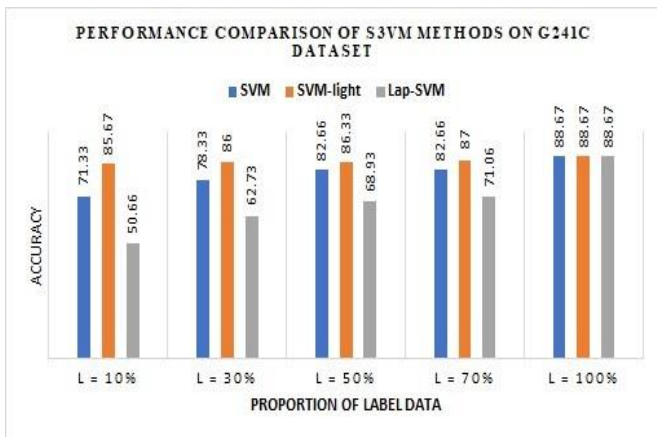


Figure 2. Performance comparison of semi-supervised SVM on G241c data set

Table 4. Performance of semi-supervised SVM models on COIL2 data set

Data set	L in %	Classification accuracy		
		SVM	S3VMlight	Lap-SVM
COIL2	10	69.00	87.67	86.26
	30	81.33	89.33	98.46
	50	86.66	98.00	98.73
	70	87.00	99.00	99.80
	100		99.06	

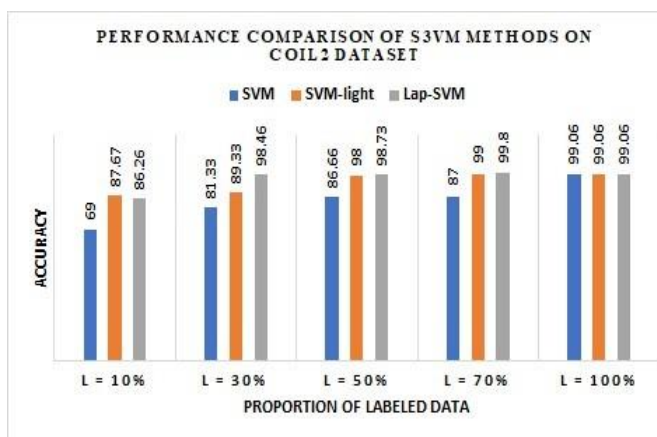


Figure 3. Performance comparison of semi-supervised SVM on COIL2 data set

Table 4 and Figure 3 show the results on COIL2 data set using all the three SSL approaches. We can observe that L=30% gives highest accuracy using lap-SVM approach.

On COIL2 data set performance of S^3VM^{light} with 50% of labeled data and Lap-SVM with 30% of labeled data is comparable with the performance of the supervised learning with complete training data labeled. This implies that semi-

supervised SVM achieves better performance with very few labeled data supplemented with large unlabeled data for training.

Based on the experimental studies we claim that semi-supervised SVM achieves better performance with very few labeled data supplemented with large unlabeled data for training.

Objective of this work is to yield better performance with limited data. Based on experimental results it can be inferred that semi-supervised version of the models that use unlabeled data yields better accuracy when compared to its counterpart supervised model.

5. CONCLUSION AND FUTURE DIRECTIONS

In this work, we present a set of semi-supervised learning techniques that can be used for various applications. Based on the experiments, we observe that semi-supervised models that uses unlabeled data while training the models improve the classification accuracy. When compared to the supervised SVM based classifiers semi-supervised SVM classifiers give better accuracy. It is observed that the Semi-supervised classifiers with few labeled data together with large amount of unlabeled data performs better than base line SVM classifiers. In future we are planning to extend to large data sets and how S3VM works for other tasks like speech and video data.

REFERENCES

- [1] Bodapati, J.D., Kishore, K.V.K., Veeranjanyulu, N. (2010), An intelligent authentication system using wavelet fusion of K-PCA, R-LDA. International Conference on Communication Control and Computing Technologies, Ramanathapuram, India. <https://doi.org/10.1109/iccct.2010.5670591>
- [2] Bodapati, J.D., Veeranjanyulu, N. (2019). Feature extraction and classification using deep convolutional neural networks. Journal of Cyber Security and Mobility, 8(2): 261-276. <https://doi.org/10.13052/jcsm2245-1439.825>
- [3] Bodapati J.D., Veeranjanyulu N. (2017) Abnormal network traffic detection using support vector data description. In: Satapathy S., Bhateja V., Udgata S., Pattnaik P. (eds), Proceedings of the 5th International Conference on Frontiers in Intelligent Computing: Theory and Applications. Advances in Intelligent Systems and Computing, vol 515. Springer, Singapore. https://doi.org/10.1007/978-981-10-3153-3_49
- [4] Nigam, K.P. (2001). Using unlabeled data to improve text classification. Carnegie-Mellon Univ. Pittsburgh Pa School of Computer Science.
- [5] Chapelle, O., Schlkopf, B., Zien, A. (2006), Semi-supervised learning MIT press. IEEE Transactions on Neural Networks, 20(3): 542-542. <https://doi.org/10.1109/TNN.2009.2015974>
- [6] Wang, Y., Chen, S.C., Zhou, Z.H. (2012). New semi-supervised classification method based on modified cluster assumption. IEEE Transactions on Neural Networks and Learning Systems, 23(5): 689-702. <https://doi.org/10.1109/TNNLS.2012.2186825>
- [7] Subramanya, A., Talukdar, P.P. (2014), Graph-Based Semi-Supervised Learning. Morgan and Claypool

- Publishers.
<https://doi.org/10.2200/S00590ED1V01Y201408AIM029>
- [8] Blum, A., Chawla, S. (2001). Learning from labeled and unlabeled data using graph min cuts. *International Conference on Machine Learning*.
- [9] Blum, A., Lafferty, J., Rwebangira, M.R., Reddy, R. (2004). Semi-supervised learning using randomized min cuts. In: *Proceedings of the Twenty-First International Conference on Machine Learning*, pp. 13. <https://doi.org/10.1145/1015330.1015429>
- [10] Joachims T. (1999). Transductive inference for text classification using support vector machines. *Proceedings of the Sixteenth International Conference on Machine Learning*, pp. 200-209.
- [11] Kveton, B., Valko, M., Rahimi, A., Huang, L. (2010). Semi-supervised learning with max-margin graph cuts. *International Conference on Artificial Intelligence and Statistics*, pp. 421-428.
- [12] Azran, A. (2007). The rendezvous algorithm: Multiclass semisupervised learning with markov random walks. In *Proceedings of the 24th international conference on Machine Learning*, pp. 49-56. <https://doi.org/10.1145/1273496.1273503>
- [13] Szummer, M., Jaakkola, T. (2002). Partially labeled classification with Markov random walks. *Advances in Neural Information Processing Systems*, pp. 945-952.
- [14] Belkin, M., Niyogi, P., Sindhwani, V. (2006). Manifold regularization: A geometric framework for learning from labeled and unlabeled examples. *The Journal of Machine Learning Research*, 7: 2399-2434.
- [15] Sindhwani, V., Niyogi, P., Belkin, M. (2005). Beyond the point cloud: from transductive to semi-supervised learning. In *Proceedings of the 22nd International Conference on Machine Learning*, pp. 824-831. <https://doi.org/10.1145/1102351.1102455>
- [16] Zhou, D.Y., Bousquet, O., Lal, T.V., Weston, J. (2004). Learning with local and global consistency. *Advances in Neural Information Processing Systems*, 16(3): 321-328.
- [17] Zhu, X., Ghahramani, Z., Lafferty, J.D. (2003). Semi-Supervised Learning Using Gaussian Fields and Harmonic Functions. In *Proceedings of the 20th International Conference on Machine learning (ICML-03)*, pp. 912-919.
- [18] Zhu, X.J., Goldberg, A.B. (2009). Introduction to semi-supervised learning. *Synthesis Lectures on Artificial Intelligence and Machine Learning*, 3(1): 1-130. <https://doi.org/10.2200/S00196ED1V01Y200906AIM006>
- [19] Bodapati, J.D., Veeranjaneyulu, N. (2019). Facial emotion recognition using deep CNN based features. *2019 International Journal of Innovative Technology and Exploring Engineering (IJITEE)*, 8(7): 1928-1931.
- [20] Joachims, T. (2003). Transductive learning via spectral graph partitioning. *Proceedings of the Twentieth International Conference on Machine Learning (ICML-2003)*, Washington DC, pp. 290-297.
- [21] Vladimir, V., Vapnik, V. (1998). *Statistical Learning Theory*. Wiley, New York.
- [22] Bennett, K., Demiriz, A. (1999). Semisupervised support vector machines. *Advances in Neural Information Processing Systems*, pp. 368-374.
- [23] Fung, G., Mangasarian, O.L. (1999). Semi-supervised support vector machines for unlabeled data classification. *Optimization Methods and Software*, 15(1): 29-44. <https://doi.org/10.1080/10556780108805809>
- [24] Chapelle, O., Sindhwani, V., Keerthi, S.S. (2006). Branch and bound for semi-supervised support vector machines. *Advances in Neural Information Processing Systems*, pp. 217-224.
- [25] Chapelle, O., Sindhwani, V., Keerthi, S.S. (2008). Optimization techniques for semi-supervised support vector machines. *Journal of Machine Learning Research*, 9: 203-233.
- [26] Garla, V., Taylor, C., Brandt, C. (2013). Semi-supervised clinical text classification with Laplacian SVMs: An application to cancer case management. *Journal of Biomedical Informatics*, 46(5): 869-875. <https://doi.org/10.1016/j.jbi.2013.06.014>
- [27] Belkin, M., Niyogi, P. (2003). Using manifold structure for partially labeled classification. *Advances in Neural Information Processing Systems*.
- [28] Melacci, S., Belkin, M. (2011). Laplacian support vector machines trained in the primal. *Journal of Machine Learning Research*, pp: 1149-1184.
- [29] Qi, Z.Q., Tian, Y.J., Shi, Y. (2014). Successive over relaxation for Laplacian support vector machine. *IEEE Transactions on Neural Networks and Learning Systems*, 26(4): 674-683. <https://doi.org/10.1109/TNNLS.2014.2320738>