Extreme Machine Learning Adaptation Under Supervision

[1]E Krishna,Assistant Professor
[2]Dr. Akilarasu Ganesan,Professor
[3]A Shiva Rama Krishana,Assistant Professor
Department of CSE Engineering,
Pallavi Engineering College Hyderabad, Telangana 501505

ABSTRACT

Agreat learning framework is manifold regularization-based semi-supervised learning (SSL). A key factor affecting SSL's performance is the way in which manifold graphs are constructed and the level of safety at which unlabeled samples are tested before encryption can begin. Unlabeled samples are often pre-constructed before classification and corrected throughout the classification learning process due to the building of several graphs and safety degrees. We propose a unified adaptive safe semi-supervised learning (Adap-SaSSL) system to address the aforesaid issues. A manifold graph is adaptively constructed and the safety degrees of unlabeled samples are adaptively calculated using this framework. As a result, the weights and parameters of the manifold graph and its parameters, as well as the safety degrees of unlabeled samples, will be optimised during the learning process rather than being precalculated. Finally, using the Adap-SaSSL architecture, we created and implemented an adaptive safe classification approach known as the adaptive safe semi-supervised extreme learning machine (ASSEM) (AdSafe-SSELM). AdSafe-effectiveness SSELM's and dependability have been shown via testing on a variety of simulated, benchmark, and image datasets.

INDEX TERMS

An adaptive graph, manifold regularisation, and a semi-supervised learning (SSL) machine (MR).

INTRODUCTION

In both theory and practise, semi-supervised learning (SSL) is an effective learning paradigm that has proven successful over the previous decade. Marker samples are typically difficult and expensive to get, yet in many practical situations it is more convenient and cost-effective to collect unmarked samples. When it comes to facial recognition [5–7], voice recognition [8–9], and handwritten digit detection, SSL is commonly used [10–12]. In general, SSL relies on a variety of assumptions, such as smooth, cluster, and manifold assumptions, to establish relationships between labelled and unlabeled samples. It is one of the most common assumptions [1] to make. Manifold regularisation techniques, such as those developed by Belkin et al. [1], may successfully use the information in unlabeled data, as shown by the results of the Laplacian regularised least squares (Lap-RLS) and Lap-SVM algorithms. It is well-known that MR's performance depends heavily on the manifold graph's architecture. As soon as a good-performing graph is built, it may eventually aid increase classification performance [10–13].. The performance will suffer as a result, and the classification will suffer as well [13]. It is also common for the graph to be established in advance and to be retained in place during the training process. It is almost hard for us to tell whether the graph is doing well before we see it. A parameter selection problem remains unsolved in semi-supervised learning with little label information because it is difficult to construct a decent performance graph before classification [13]. It adds even another hurdle to the process of creating MR graphs in advance. Few studies have focused on graph building to date [10]–[13], so far as we know, current MR enhancements either strive to determine the regularisation parameters or increase the efficiency of MR. SSL performance may be negatively impacted by the use of unlabeled samples, according to recent research [14], [15]. SSL's practical applicability will be limited if unlabeled samples are not safe to utilise [14]. A safe semi-supervised learning (SaSSL) approach is thus required, since the SL technique utilising just labelled samples is always poorer than the SaSSL approach. [16]–[20] and [20]–[22] are examples of semi-supervised learning approaches that have been proposed in recent years. ELM is a new SLFNs technique suggested by Huang et al. [23] as a single hidden layer feed-forward network. Some ELM-based alternative algorithms have recently been suggested as a result of the good performance of ELM. Because they are virtually all supervised learning algorithms, they are unable to make use of unlabeled examples in training their models. MR-based semi-supervised ELM method by Huang et al. [30] has been presented as a solution to this issue. Unfortunately, SS-ELM currently lacks the required safe mechanisms when attempting to exploit unlabeled samples. When using the SS-ELM with unlabeled samples, the results may be less accurate. SHE et al. [31] presented a safe semi-supervised extreme learning machine in answer to the aforementioned issue (Safe SSELM). When just labelled samples are used in the experiments, Safe-performance SSELM's is seldom worse than that of ELM. Adaptive safety semi-supervised learning (Adsfe-SSL) is proposed in this research. In the first step, we build a manifold graph by combining sparse restrictions with classification learning. Instead than having to be pre-specified, the manifold and its parameters may be automatically modified as the student learns new material. Entropy constrained adaptation is used to determine the safety of each unlabeled sample in the second step of our procedure. The Adsfe-SSL framework is the foundation for our new, adaptive, safe, semi-supervised, extreme learning machine (Adsafe-SSELM). The Adsafe-SSELM problem is also solved using the alternating iterative technique. Because each iteration yields a closed-form answer, iterative convergence may be safely assumed. The suggested technique has been shown to be successful and dependable in comparison to previous algorithms in this field by testing on different data sets. The following are the paper's primary contributions: A unified Adsfe-SSL (adaptive safe semi-supervised learning) framework is devised. On the basis of the Adsfe SSL architecture, we present a new adaptive safe semisupervised extreme learning machine (Adsafe-SSELM). (3) In our algorithm's iterative solution, we adaptively calculate the safety degrees of unlabeled samples and the construction of a manifold network. When compared to other algorithms in the same field, our method performs very well. There are three sections to this study. Our discussion of related work in Section II includes material on ELM and SS-ELM. In Section III, we'll explain our method in more depth. Section IV presents the findings and conclusions of the experiments. Section V concludes the article, and future research is discussed.

BACKGROUND

The ELM [23]–[25] and SS-ELM [30] will be briefly discussed in this section.

A. ELM In the field of machine learning and pattern recognition, the Extreme Learning Machine (ELM) is a powerful instrument. ELM produces the hidden layer's input weights and biases at random. As a result, ELMs have the benefits of a simple structure, a cheap computational cost, and a high degree of flexibility over typical neural networks. The limitations of classic neural networks such as local minima, inaccurate learning rates and poor convergence rates are solved by ELMs. There are n training samples in the training set, and the number of samples in the training set is l. The output function of ELM [23]–[25] is given by L if L is the number of neurons in the hidden layer.

$$\boldsymbol{Y} = \boldsymbol{H}\boldsymbol{\beta}$$

where [1, 2,.., L] is the number of variables. the hidden layer of L nodes, Y = [y1, y2,..] and the output node, T is the vector of output weights between these nodes T, H is the output matrix of the hidden layer.

$$\boldsymbol{H} = \begin{bmatrix} h_1(x_1) & \cdots & h_L(x_1) \\ \vdots & \vdots & \vdots \\ h_1(x_l) & \cdots & h_L(x_l) \end{bmatrix}$$

which has hi(x) = G(ai, bi, x) = (ai + bi) (x + I = 1,...L where (a and b can be randomly generated according to a continuous probability distribution). A simple way to explain the ELM framework's primary regularisation is:

$$\min_{\beta} \Psi(\beta) = C\|\boldsymbol{H}\beta - \boldsymbol{Y}\|^2 + \|\beta\|^2$$

where C is the training error penalty coefficient. According to the Moore-Penrose principle, the output weight vector is then generated. If l is less than L, then the answer to (2) is

$$\boldsymbol{\beta}^* = (\boldsymbol{H}^T\boldsymbol{H} + \frac{\boldsymbol{I}_L}{C})^{-1}\boldsymbol{H}^T\boldsymbol{Y}.$$

where IL is an identity matrix of dimension L. If l < L, the solution of (2) is:

$$\boldsymbol{\beta}^* = \boldsymbol{H}^T(\boldsymbol{H}\boldsymbol{H}^T + \frac{\boldsymbol{I}_l}{C})^{-1}\boldsymbol{Y}.$$

In this case, Il is an identity matrix of dimension l B. SS-ELM It is true that many fields employ ELMs, but their primary usage is in supervising learning tasks like classification and regression, which severely restricts its usefulness In reality, however, obtaining labelled samples is often difficult and costly, while obtaining vast volumes of unlabeled samples is simple and affordable [1]. Semi-supervised extreme learning machine (SS-ELM) method was suggested by Huang et al. [30] in order to solve the problem of supervised ELMs learning algorithms that they cannot employ unlabeled data. Consider using an SSL with a training set. If xi, yi and u are all in the same range, then T is equal to Tl, and Tu is equal to Tl+u, where T is equal to Tl+u and Tu is equal to Tuxi, which is equal to Tlxy, where T is equal to Tlxy and Tuxy, where T is equal to Tuxy. There are l labelled samples, and u unlabeled samples in the Tl and Tu datasets. SS-ELM [30]'s primary difficulty may therefore be stated as:

In this case, L = D W is referred to as the graph Laplacian since it incorporates data from both labelled and unlabeled samples. D is a diagonal matrix, while W is the similarity matrix.

III. ADAPTIVE SAFE SEMI-SUPERVISED EXTREME LEARNING MACHINE

MOTIVATION

In this part, we'll go through the fundamentals of our algorithm in great depth. An adaptive safe method was created to determine the security of each unlabeled sample and assign different safe degrees to various unlabeled samples.. Unlabeled samples are safer than labelled samples with risk because of the safe-based trade-off between SL and SSL, which assures that unlabeled samples are safer. The entropy maximisation criteria was used to attain the aforementioned objectives. Second, we provide a system for automatically adapting the graph's construction and parameters as the learner progresses, rather than requiring prior reservation. Consequently, this adaptive graph building and classification learning provides a powerful combination. B. INNOVATIVE RISK DENSITY AND GRAPH It is called f (x) and g (x) for the semi-supervised and supervised classifiers. A simplistic solution is avoided by using an Entropy Maximization criteria [6,10], which requires a uniform distribution of the unlabeled samples' safety degrees. As a result, the phrase "adaptive safety degree" may be defined as follows:

$$\begin{aligned}
\min_{s_j} \quad & R(s_i) = \sum_{j=l+1}^{n} s_j \|f(x_j) - g(x_j)\|^2 \\
& \qquad + \sum_{j=l+1}^{n} s_j \ln(s_j) \\
\text{s.t.} \quad & \sum_{j=l+1}^{n} s_j = 1, \\
& 0 \le s_i \le 1, \quad \forall i = l+1, \dots, n \qquad (8)
\end{aligned}$$

where sj is a measure of the safety of samples xj that have not been labelled. We present an adaptive graph mechanism that introduces sparse restrictions in order to better optimise graph creation and parameter modification during graph building. Because each sample in the manifold graph is only related to a few other samples, the weights of each weight should only include a few non-zero elements and the rest should all be zeros. Wi weight vectors should all be very small. Consequently, the term adaptive graph may be summarised as:

ADAPTIVE SAFE SEMI-SUPERVISED LEARNING

Based on (8) and (9) we can get a unified adaptive safe semisupervised learning (Adap-SaSSL) framework

$$\begin{aligned} \min_{f, w_{i,j}, s_j} \quad & \sum_{i=1}^{l} (f(x_i) - y_j)^2 + \lambda_1 \|f\|_{\mathcal{H}}^2 \\ & + \lambda_3 \Xi(w_{i,j}) + \lambda_2 R(s_i) \\ \text{s.t.} \quad & \sum_{j=l+1}^{n} s_j = 1, \\ & 0 \le s_i \le 1, \quad \forall i = l+1, \dots, n \\ & \sum_{j=1}^{u} w_{ij} = 1, \\ & 0 \le w_{ij} \le 1, \quad \forall i = 1, \dots, n \end{aligned}$$

where the regularisation parameters are 1, 2 and 3, the safety level of unlabeled samples xj is described by si, the graph weights are constrained by 4(wi,j), and the risk level of unlabeled samples is constrained by R(si). This framework uses a manifold network and its parameters that are optimised in the learning process rather than pre-defined, and safety degrees for distinct unlabeled samples are calculated and assigned adaptively. It's possible to create an adaptive safe semi-supervised ELM learning framework based on the adaptive safe semi-supervised ELM (AdSafe-SSELM) framework (10).

$$\begin{aligned} \min_{\beta, w_{i,j}, s_j} \quad & \sum_{i=1}^{l} ((h(x_i)\beta - y_j)^2 + \|\beta\|^2 \\ & + \lambda_1 \Xi(w_{i,j}) + \lambda_2 R(s_i) \\ \text{s.t.} \quad & \sum_{j=l+1}^{n} s_j = 1, \\ & 0 \le s_i \le 1, \quad \forall i = l+1, \dots, n \\ & \sum_{j=1}^{u} w_{ij} = 1, \\ & 0 \le w_{ij} \le 1, \quad \forall i = 1, \dots, n \end{aligned}$$

In order to solve the non-convex optimization issue of AdSa-SSELM, we will use the alternate iterative technique to get the output weights, graph weights wi,j, and the risk of unlabeled samples sj. Fortunately, a closed-form solution exists for each step. Step 1: We can achieve secure semi-supervised ELM (Safe-SSELM) with fixed wi, j, and sj [31]

D. ANALYSIS OF TIME COMPLEXITY

With regard to the number of labelled samples and hidden neurons, ELM has the same computational complexity as L-ELM, which has the computational complexity of O(l + u) 3 + L(l + u) 2 (l denotes the number of labelled samples and u the number of unlabeled data). For Algorithm 1, the number of iterations per iteration and the computing cost define the computational complexity. In the beginning, we analysed the other option. This algorithm has three main parts: (1) The solution of , which is similar to that in SS-ELM, has an O((l + u) 3 + L(l + u) 2 complexity; (2) Calculating the computational complexity of (sj) through (21) is equivalent to O((l + u) 2 log(l + u) 2; and (3) Calculating the computational complexity of (wki) through (30) is equivalent to O(l + u) 2. As a result, Algorithm 1 has a computational complexity of T O((l + u) 2 log(l + u) + (l + u) 3 + (L + 1)(l + u) 2). When it comes to the iterative number T, our research shows that T = 100 comes close to being ideal.

V. CONCLUSIONS

Adaptive safe semi-supervised learning (AdSafe-SSL) has been suggested in this research. On top of the manifold graph of learning, this method also calculates the adaptive safety level for each unlabeled sample. We've developed an AdSafe-SSELM-based adaptive safe semi-supervised extreme learning system. Results from multiple datasets show that AdSafe-performance SSELM's is never significantly worse than ELM's or SS's, indicating the efficacy of our safe mechanism design. In the future, we'll look at different methods for assessing the danger of using unlabeled samples, as well as various variations on the semi-supervised learning safety mechanism.

REFERENCES

[1] M. Belkin, P. Niyogi, and V. Sindhwani, ''Manifold regularization: A geometric framework for learning from labeled and unlabeled examples,'' J. Mach. Learn. Res., vol. 7, pp. 2399–2434, Nov. 2006.

[2] O. Chapelle, B. Schlkopf, and A. Zien, ''Semi-supervised learning,'' in Handbook on Neural Information Processing. Berlin, Germany: Springer, 2013.

[3] P. K. Mallapragada, R. Jin, A. K. Jain, and Y. Liu, ''SemiBoost: Boosting for semi-supervised learning,'' IEEE Trans. Pattern Anal. Mach. Intell., vol. 31, no. 11, pp. 2000–2014, Nov. 2009.

[4] X. Zhu, ''Semi-supervised learning literature survey,'' Comput. Sci., Univ. Wisconsin-Madison, Tech. Rep., 2008.

[5] H. Gan, N. Sang, and R. Huang, ''Self-training-based face recognition using semi-supervised linear discriminant analysis and affinity propagation,'' J. Opt. Soc. Amer. A, Opt. Image Sci., vol. 31, no. 1, pp. 1–6, 2014.

[6] F. Roli and G. L. Marcialis, ''Semi-supervised PCA-based face recognition using self-training,'' in Proc. Joint IAPR Int. Workshops Struct., Syntactic, Stat. Pattern Recognit. (SSPR), vol. 4109, Hong Kong, Aug. 2006, pp. 560–568.

[7] Y. Cao, H. He, and H. Huang, ''LIFT: A new framework of learning from testing data for face recognition,'' Neurocomputing, vol. 74, no. 6, pp. 916–929, 2011.

[8] D. Yu, B. Varadarajan, L. Deng, and A. Acero, ''Active learning and semisupervised learning for speech recognition: A unified framework using the global entropy reduction maximization criterion,'' Comput. Speech, Lang., vol. 24, no. 3, pp. 433–444, 2010.

[9] S. Van Vaerenbergh, I. Santamaria, and P. E. Barbano, ''Semi-supervised handwritten digit recognition using very few labeled data,'' in Proc. IEEE Int. Conf. Acoust., Speech Signal Process., vol. 7882, May 2011, pp. 2136–2139.

[10] B. Geng, D. Tao, C. Xu, L. Yang, and X.-S. Hua, ''Ensemble manifold regularization,'' IEEE Trans. Pattern Anal. Mach. Intell., vol. 34, no. 6, pp. 1227–1233, Jun. 2012.

[11] K. Zhang, J. T. Kwok, and B. Parvin, ''Prototype vector machine for large scale semi-supervised learning,'' in Proc. Int. Conf. Mach. Learn., 2009, pp. 1233–1240.

[12] I. W. Tsang and J. T. Kwok, ''Large-scale sparsified manifold regularization,'' in Proc. Adv. Neural Inf. Process. Syst., 2007, pp. 1401–1408.

[13] Y. Wang, Y. Meng, Y. Li, S. Chen, Z. Fu, and H. Xue, ''Semi-supervised manifold regularization with adaptive graph construction,'' Pattern Recognit. Lett., vol. 98, pp. 90–95, Oct. 2017.

[14] Y.-F. Li and Z.-H. Zhou, ''Improving semi-supervised support vector machines through unlabeled instances selection,'' in Proc. AAAI Conf. Artif. Intell., AAAI Press, 2011, pp. 386–391.

[15] Y.-F. Li and Z.-H. Zhou, ''Towards making unlabeled data never hurt,'' in Proc. Int. Conf. Mach. Learn., Omnipress, vol. 37, 2011, pp. 1081–1088.

[16] Y. Wang and S. Chen, ''Safety-aware semi-supervised classification,'' IEEE Trans. Neural Netw. Learn. Syst., vol. 24, no. 11, pp. 1763–1772, Nov. 2013.

[17] M. Kawakita and J. Takeuchi, ''Safe semi-supervised learning based on weighted likelihood,'' Neural Netw., vol. 53, no. 5, pp. 146–164, 2014.

[18] H. Gan, Z. Z. Luo, M. Meng, Y. Ma, and Q. She, ''A risk degree-based safe semi-supervised learning algorithm,'' Int. J. Mach. Learn., vol. 7, no. 1, pp. 85–94, 2015.

[19] H. Gan, Z. Luo, Y. Sun, X. Xi, N. Sang, and R. Huang, ''Towards designing risk-based safe Laplacian regularized least squares,'' Expert Syst. Appl., vol. 45, pp. 1–7, Mar. 2016.

[20] H. Gan, Z. Li, Y. Fan, and Z. Luo, ''Dual learning-based safe semisupervised learning,'' IEEE Access, vol. 6, pp. 2615–2621, 2017.