International Journal on Emerging Technologies 11(2): 283-288(2020)
ISSN No. (Print): 0975-8364
ISSN No. (Online): 2249-3255

Adaptive Cost-Sensitive Sparse Representation Truncated Gradient Online Classification (ACSSRTGC)

P. Priakanth[1], M. Thangamani[2] and S. Gopikrishnan[3]
[1]Professor & Head, School of Computer Technology and Applications,
Kongu Engineering College, Perundurai (Tamil Nadu), India.
[2]Assistant Professor, Department of Computer Science and Engineering,
Kongu Engineering College, 638060 Perundurai (Tamil Nadu), India.
[3]Assistant Professor, School of Computer Science and Engineering,
VIT-AP University, Amaravati 522237 (Andhra Pradesh), India.

(Corresponding author: P. Priakanth)
(Received 04 December 2019, Revised 30 January 2020, Accepted 05 February 2020)
(Published by Research Trend, Website: www.researchtrend.net)

ABSTRACT: The data stream classification tasks are batch learning, suffers from poor efficiency and lack of scalability. Recently cost-sensitive online classification algorithms pooled with adaptive regularization introduced classification. In practical applications, most of the data set is imbalanced state in expressions of class division. The proposed Adaptive Cost-Sensitive Sparse Representation Truncated Gradient based Classification (ACSSRTGC) for the difficulty of class-imbalance. The test samples, class labels are predicted by reducing the losses incurred due to misclassification, acquired through the computation of the posterior probabilities. The proposed algorithms are theoretically analysed and their efficiency and characteristics are empirically validated in elaborate experiments via the German and Covtype dataset. The results of the methods are measured via the classification metrics and implemented in MATLAB environment. This research experiment result outperforms the cost sensitive online gradient descent technique and adoptive regularized cost-sensitive online descent approach.

Keywords: Adaptive Regularization, Sketching Learning, Online Learning, Cost-Sensitive Classification, Truncated Gradient algorithm and Sparse Representation.

I. INTRODUCTION

Spare online learning and cost sensitive learning focuses on new convex optimization. Empowered with the exponential rise in datasets, data mining and machine learning technologies given a great support to the modern time society in several aspects ranging from filtering the content to web search performed on social sites, and from products' recommendations to intelligent consumer services via e-commerce. Usually, several realistic large-scale applications employ numerous techniques referred as online learning. It was elaborately researched for several years in data mining and machine learning domains [1-2]. Usually, online learning involves a group of useful as well as machine learning approaches which are scalable, whose intent is the incremental learning of a model for performs the predictions with accuracy on a pool of samples. Class imbalance problems are represented [3-5].

Online learning is noteworthy due its superior scalability and efficiency achieved in applications of large-scale. It is utilized for finding a resolution for classification tasks online in several data mining applications in practice. In spite of being researched deeply in machine learning, many of the presented online learning methodologies are adverse and probably would not prove to be sufficiently useful for finding a solution for cost-sensitive classification task, which is a major data mining problem that considers the costs of misclassification. Because the fact that many of the available online learning research are mostly associated with the performance achieved out of an online classification algorithm with respect to accuracy rate or prediction error rate, which is in fact insensitive to cost and therefore unsuitable for several practical applications in the field of data mining, particularly for tasks involving cost-critical classification in which the datasets are frequently in imbalanced state with respect to class. Misclassification cost of instances from various classes could be relatively dissimilar [6].

Consequently, Cost-Sensitive Online Classification framework [7-8] is introduced some time before for filling gap in cost critical classification and online learning. In this, algorithms group called Cost-Sensitive Online Gradient Descend (COG) is suggested for direct optimization of predetermined cost-based metrics depending on online gradient descent method. In order to improve COG algorithm, Adaptive Regularized Cost-Sensitive Online Gradient Descent algorithms (ACOG), which is grounded on benchmarked Confidence Weighted mechanism is planned. In practical, imbalance in the class distribution is observed in most of the data sets. In lieu with further investigation on the deep theory of cost-critical online learning, it is highly suggested to analyse the sparse computation techniques involved in problems involving cost sensitive online classification. With this motivation, investigated an Adaptive Cost-Sensitive Sparse Representation Truncated Gradient based classification (ACSSRTGC) for the problem of class-imbalance. Next, elaborate experiments are carried out to assess the performance and efficiency of the newly introduced algorithms and then they are applied for finding a resolution to the problem involving online anomaly detection tasks from various practical research fields. Potential results help in reassuring the efficiency and usefulness of the techniques in practical cost-critical online classification issues.

II. LITERATURE REVIEW

Wang *et al.,* recommended a new design aimed at cost-sensitive online classification that uses ideas of online gradient descent strategies. According to the design, these algorithms are used for optimizing two prevalent cost-sensitive facets: specificity and sensitivity weighted summation increase and expenses of weighted misclassification reduction [7].

Lee and Stephen briefs the iterative methods to calculate the subgradient directions which is useful to stochastic problems over large and streaming data set. [8].

Crammer *et al.*, proposed a set of margin based online learning algorithms developed for different tasks of prediction. Specifically algorithms are formulated and evaluated with the objective of resolving binary and multiclass classification, regression, uniclass prediction and sequence prediction. The steps on updating different algorithms all rely on analytical solutions for rather ordinary constraint based optimization problems. This amalgamated view helps in showing the worst-case loss thresholds for various algorithms and for different decision problems that depends on one individual lemma. The thresholds on the aggregated loss of the algorithms are associated with the least loss, which can be achieved with the help of any static hypothesis, and like it is, they are suitable for both implementable and non-implementable configuration [9].

It emphasis the merits of two learning paradigms are, it incrementally update the trained model, measure and sparsity of solutions [10]. Zhenbing Liu *et al.*, briefs about how to reduce the cost of misclassification loses in cost sensitive approach [11]. Ma compares with the original truncated variants and prediction accuracy to find the better result [12]. The approach enables us to handle the regular and high dimensional spares data for more effective learning done by [13].

Zhang *et al.*, proposed a novel Streaming Features algorithm (OL_{SF}) including Online Learning in addition with two modifications, integrating selection of streaming features and online learning for aiding trapezoidal data sets learning filled with huge number of training features and examples. In case another new training instance with novel features springs up, classifier sets on updating the features, which exist already by using passive-aggressive update rule. Structural risk reduction principle is obtaining updated new features. Sparsely of feature is proposed employing projected truncation approach [14].

Yan *et al.,* investigated a technique known as Online Heterogeneous Transfer (OHT) which utilize hedge ensemble employing both offline and online data on different domains. For this purpose, formulated data of unlabeled auxiliary co-occurrence. Labelled source data is used to design an offline decision function grounded on a heterogeneous similarity. Then target data is employed for online decision function learning. At last, a hedge weighting strategy is utilized for combining decision functions of online and offline in order to use information obtained from multiple feature space's target and source domains. Moreover, conventional evaluation on error bounds of novel technique is provided. Elaborate experiments carried out on three practical data sets prove the efficiency of the novel approach [15].

Zhao *et al.,* introduced a cost-sensitive online classification algorithm group. Adaptive regularization is combined with it and implemented. The newly introduced algorithms are theoretically evaluated and then empirically validated using elaborate efficiency to prove their efficiency and characteristics. Thereafter, to achieve a better balance between the performance and efficacy, further sketching approach is presented into the algorithms, which considerably improves the speed of computation with negligible loss in performance. At last, the algorithms are used for dealing with various tasks of online anomaly detection found in practice. The results achieved indicate that the newly introduced algorithms are useful and effective in finding a resolution to problems involving cost-sensitive online classification in a variety of practical fields [16-20].

III. PROPOSED METHODOLOGY

Adaptive Cost-Sensitive Sparse Representation Truncated Gradient based classification (ACSSRTGC) is proposed for the problem of class-imbalance. The class label of test samples are predicted by reducing the losses of misclassification, acquired through the posterior probabilities' computation. For L_1-regularization, ordinary online sub-gradient method is combined with end of training rounding. Parameter of regularization is used in this. Figure 1 illustrates the flow diagram of the novel system.

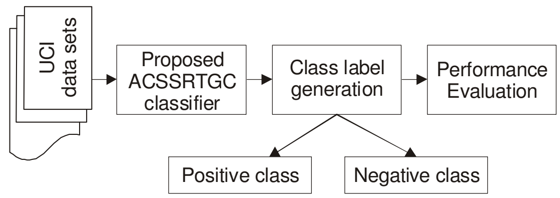

Fig. 1. Flow chart of proposed ACSSRTGC system.

A. Problem Setting

The important objective is the learning of a linear classification model that uses an updatable predictive vector$w \in \mathbb{R}^d$, on a set of training samples $\{(x_1, y_1), \dots.(x_T, y_T)\}$, where T refers to the overall number of samples, $x_t \in \mathbb{R}^d$ stands for sample with d-dimension at time t, and $y_t \in \{1, -1\}$stands for correspondinglabelof true class. To understand, during t^{th} learning round, learner receives a sample x_t and class label $\hat{y}_t = \mathrm{sign}(w_t^T x_t)$, is predicted where w_t refers to predictive vector of model, which is learnt from previous$t-1$ samples. After that, model receives instance base truth$y_t \in \{1, -1\}$, which, in turn is taken as true class label. In case $\hat{y}_t = y_t$, the model has predicted correctly; else, there is an error and hence forth a loss is incurred. Finally, the learner updates its predictive vector w_t depending on the incurred distressful loss. The newly introduced ACOG algorithms are presented by the objective optimization. But, non-convex property is exhibited by objective function. In order to solve this optimization, indicator function is used with its variants of convex. They are described as,

$$\ell^{I}\left(w:(x,y)\right) = \max\left(0, \left(\rho * \mathbb{I}_{(y=1)} + \mathbb{I}_{(y=-1)} - y(w.x)\right)\right) \quad (1)$$

$$\ell^{II}\left(w:(x,y)\right) = \left(\rho * \mathbb{I}_{(y=1)} + \mathbb{I}_{(y=-1)} * \max\left(0,1 - y(w.x)\right)\right) \quad (2)$$

For $\ell^{I}\left(w:(x,y)\right)$, the change observed in the margin provides more "frequent" updates for a particular class, in comparison with the classical hinge loss; whereas for $\ell^{II}\left(w:(x,y)\right)$, change in slope results in highly "aggressive" updates of particular class. Therefore, objective is to reduce learning remorse procedure [14], depending on either of the loss functions $\ell^{I}\left(w:(x,y)\right)$ or $\ell^{II}\left(w:(x,y)\right)$:

$$Regret := \sum_{t=1}^{T} \ell\left(w_t;(x_t,y_t)\right) - \sum_{t=1}^{T} \ell\left(w^*;(x_t,y_t)\right) \quad (3)$$

Where

$w^* = \arg\min_t \sum_{t=1}^{T} \nabla\ell(w; (x_t, y_t))$ (4)

In order to obtain a resolution for this problem of optimization, Cost-sensitive Online Gradient descent algorithms (COG) [4-5] were presented:

$w_{t+1} = w_t - \eta\nabla\ell_t(w_t)$ (5)

Where ηrefers to the learning rate and

$\ell_t(w_t) = \ell(w; (x_t, y_t))$ (6)

IV. COG ALGORITHM

COG algorithms takes the sample set's first order gradient information only into consideration for updating the learner, which in turn, is quite inadequate as several studies in the recent times have proven the importance of including the second order information [6, 9]. Inspired by this finding, adaptive regularization is proposed to be introduced to encourage the cost-sensitive online categorization. Let it be presumed that online model fulfils multivariate Gaussian distribution, i.e., $\sim \mathcal{N}(\mu; \sum)$, where μrefers to distribution's mean value vector and $\sum$specifies distribution's covariance matrix. After this, the class label of a sample x can be predicted on the basis of sign (w>x), if specified definite multivariate Gaussian distribution. Practically, it is more feasible to predict by just employing distribution mean $\mathbb{E}[w] = \mu$ instead of w. Therefore, the rule of model prediction, in reality, uses sign(μ^Tx) below. To understand better, everyaverage value μ_i can be regarded to be theknowledge of model has about feature i; and covariance matrix's diagonal entry $\Sigma_{i,i}$ is considered to be feature i's confidence. Usually, lesser value of $\Sigma_{i,i}$, themore is confidence in average weight μ_i for feature 'i'. Along with diagonal values, the remaining covariance terms $\Sigma_{i,j}$ can be considered to be the correlations between two average weight value μ_i and μ_{i_j} for feature i and j.Unconstrained objectives are reduced for recasting object functions. It uses multivariate Gaussian distribution. Based on difference obtained between probability and empirical distribution, reduction of unconstrained is done as,

$D_{KL}(\mathcal{N}(\mu, \Sigma)||\mathcal{N}(\mu_t, \Sigma_t)) + \eta\ell_t(\mu) + \frac{1}{2\gamma} x_t^T \sum_t x_t$ (7)

WhereD_{KL}refers to the Kullback- Leibler Divergence (KLD), ηindicates the fitting parameter and γ stands for the regularized parameter. Particularly, this goal aids in attaining the balance between distribution divergence (first component), loss function (next component) and model confidence (final component). In other words, objective is certainly attempt doing least degree of adjustment at every round for reducing loss as well as achieving confidence optimization of model. This optimization problem can be resolved by first depicting the KLD in an explicit manner:

$D_{KL}(\mathcal{N}(\mu, \Sigma)||\mathcal{N}(\mu_t, \Sigma_t)) = \frac{1}{2}\log\left(\frac{det\Sigma_t}{det\Sigma}\right) + \frac{1}{2}\mathrm{Tr}(\sum_t^{-1}\Sigma) + \frac{1}{2}||\mu_t - \mu||^2_{\Sigma_t^{-1}1} - \frac{d}{2}$ (8)

One easy technique used for resolving objective function is to form two sections based on μ and Σ, respectively. In next step, mean vector μ update and covariance matrix Σ is performed one by one:

Parameter of mean is updated as follows:

$\mu_{t+1} = \arg\min_\mu f_t(\mu, \Sigma)$ (9)

If $\ell_t(\mu_t) \neq 0$, the covariance matrix is updated

$\Sigma_{t+1} = \arg\min_\Sigma f_t(\mu, \Sigma)$ (10)

The important concept behind SACOG is the approximation of the second covariance matrix with less number of cautiously chosen directions, known as a sketch. The improved variant of ACOG is presented through Oja's sketch technique [15], which is developed to improve the computational efficiency if the second order matrix of serial data is of lower rank. For convenience, define $\mathcal{M} = \{t|y_t \neq \mathrm{sign}(w_t.x_t), \forall t \in \{T\}\}$stands for the mistake index set, $\mathcal{M}_p = t \in \mathcal{M}$and $y_t = 1$ indicates the positive set of mistake index and$\mathcal{M}_n = \{t \in \mathcal{M}, y_t = -1\}$and $y_t = -1$ refers to the negative set of mistake index. In addition, set $\mathcal{M} = |\mathcal{M}|, \mathcal{M}_p = |\mathcal{M}_p|$ and $\mathcal{M}_n = |\mathcal{M}_n|$towards indicates the number of overall errors, positive errors and negative errors. Also, $\mathcal{I}_T^p = \{i \in [T]|y_i = +1\}$ and $\mathcal{I}_T^n = \{i \in [T]|y_i = -1\}$represents negative and positive samples index set, where $T_p = |\mathcal{I}_T^p|$ represents number of positive samples and $T_n = |\mathcal{I}_T^n|$ represents number of negative samples. For measuring this problem's performance metrics, positive samples are first assumed to be rare class, i.e., $T_p \leq T_n$.

A. Truncated Gradient

Forgetting an online variant of simple rounding rule given in expression (4), it is noted that direct rounding to zero is extremely aggravated. Much small aggravated variant helps to reduce coefficient to zero by much lesser amount. This concept is the truncated gradient. Shrinkage level is computed with help of a gravity parameter g_i> 0:

$f(w_i) = T_1(w_i - \eta\nabla_1 L(w_i, z_i), \eta g_i, \theta)$ (11)

where, v = [v_1,...,v_d]∈R^d is a vector and g ≥ 0, T_1(v,α,θ)= [T_1(v_1,α,θ),...,T_1(v_d,α,θ)] is a scalar with $T_1$and it is given by,

$$T_1(v_j, \alpha, \theta) = \begin{cases} \max(0, v_j - \alpha) if v_j \in [0, \theta] \\ \min(0, v_j + \alpha) if v_j \in [-\theta, 0] \\ v_j \text{ else} \end{cases} \quad (12)$$

Also, the truncation operation can be carried out for every K online step. This implies that, integer is not produced by case i/K, assume g_i= 0; ifi/K is an integer, assume g_i = Kg to a gravity parameter g > 0. Exact selection equals to (4) set g so that, ηKg ≥ θ. If η is small, large value of g is required. Basically, a small, constant g has to be set, as indicated by regret threshold developed at a further process. Generally, bigger parameters g and θ are, more amount of sparseness is inflicted. Owing to additional truncation T_1, this technique can result in solutions which are sparse, as shown by experiments explained as follows. In these experiments, the level of sparseness found differs with problem. One specific case, which will be tried in experiment, is to fix g = θ. Here, just one parameter g is used for controlling sparseness. As ηKg≪ θ when ηK is a smaller value, truncation operation gets small aggravated compared to rounding operation in (4). First, procedure is found to be a random means of fixing (4). Fixing θ = ∞ gives one significant and special case, which is expressed as

$f(w_i) = T(w_i - \eta\nabla_1 L(w_i, z_i), g_i\eta)$ (13)

where, v = [v_1,...,v_d] ∈ R^d is a vector and g ≥ 0, T(v, α) = [T(v_1,α),...,T(v_d, α)] is a scalar, with

$$T(v_j, \alpha) = \begin{cases} \max(0, v_j - \alpha) if v_j > 0 \\ \min(0, v_j + \alpha) \text{else} \end{cases} \quad (14)$$

The technique is an improved version of conventional sub-gradient descent technique provisioned by L1-regularization. Parameter g_i ≥ 0 regulates sparseness, which the algorithm can accomplish. It is to be noted that if g_i = 0, standardized random gradient descent and update rule are similar. The reason behind this (in place of a steady g) is that a more persuasive truncation can be performed using gravity parameter Kg after every K step. This result in sparseness that is improved.

V. EXPERIMENTS

Performance and features of existing algorithms such as ACOG, COG and proposed ACSSRTGC algorithm are evaluated in this section. Then, the effectiveness and resourcefulness of sketched versions are further evaluated. On each dataset that is German and Covtype, the experiments were carried out on arbitrary permutations of instances. The results are demonstrated in terms of the performance in average of 20 runs and assessed in terms of three metrics, which include specificity, sensitivity and weighted summation of specificity and sensitivity. Table 1 illustrates details of dataset. The potential results help in showing that the newly introduced algorithms are quite useful and effective in finding a resolution to problems in cost-sensitive online classification in diverse real-time fields.

Table 1: Details about dataset.

Dataset	No. of examples	No. of features	No. of positive	No. of negative
German	1000	24	1	2.3
Covtype	581012	54	1	1

A. Evaluation with Sum Metrics

A more apt strategy is to compute aggregated sum of weighted specificity and sensitivity:

$$\text{sum} = \alpha_p \times \frac{T_p - \mathcal{M}_p}{T_p} + \alpha_n \times \frac{T_n - \mathcal{M}_n}{T_n} \quad (15)$$

where $\alpha_p, \alpha_n \in$[0; 1] refer to the weight parameters for a balancebetween sensitivity and specificity, and $\alpha_p + \alpha_n = 1$.

It is to be observed that, if p = n = 0:5, then sum metric tends to become the well-known metric of accuracy that is balanced. Moreover, one more metric that needs to be measured is the misclassification cost that the model has to incur:

$$\text{cost} = c_p * \mathcal{M}_p + c_n * \mathcal{M}_n \quad (16)$$

where $c_p, c_n \in$ [0; 1] refer to the cost parameters ofmisclassification intended for positive as well as negative instances, $c_p + c_n$= 1. Generally, higher sum value, better would be the classification performance. Table 2 provides the summary of the experimental results achieved of the three classifiers on two datasets in terms of sum, sensitivity and specificity, and Fig. 2-4 shows the evolution of online cost performance during every iteration.

Table 2: Assessment of the classification methods and metrics.

Dataset Name	Metrics	Methods		
		COG	**ACOG**	**ACSSRTGC**
Covtype	Sum (%)	53.846	68.154	71.55
	Sensitivity (%)	50.677	72.638	74.250
	Specificity (%)	57.015	63.670	68.850
German	Sum (%)	55.998	65.155	68.31
	Sensitivity (%)	40.286	53.92	56.50
	Specificity (%)	71. 71	76.39	80.12

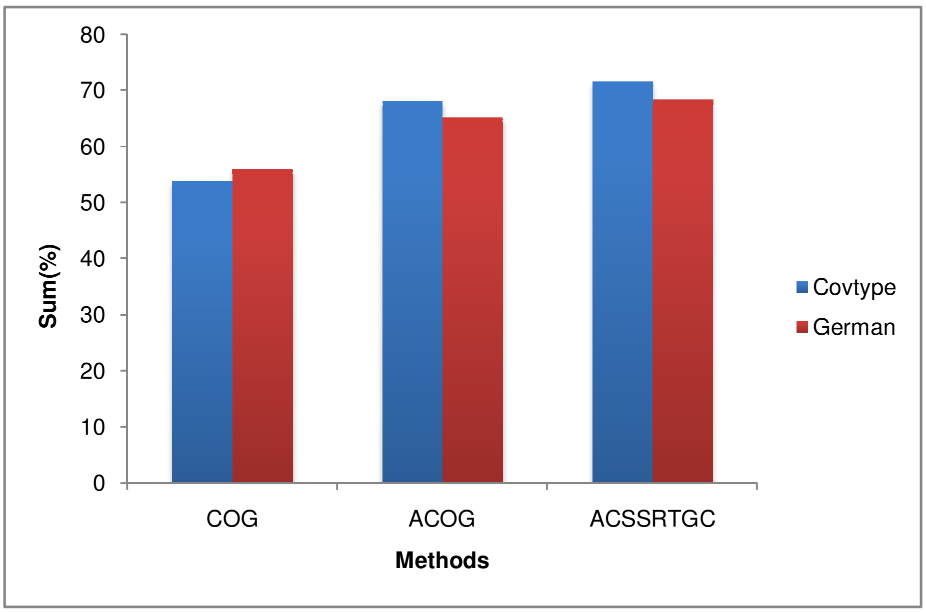

Fig. 2. Sum Results Evaluation of classification methods.

Fig. 2 shows performance comparison results of sum metric with respect to three different classifiers such as COG, ACOG and ACSSRTGC on two datasets such as Covtype and German. The proposed ACSSRTGC algorithm gives higher sum rate of 71.55% for Covtype dataset, whereas other methods such as COG and ACOG gives only 53.84% and 68.15% respectively.

Fig. 3 shows performance comparison results of sensitivity metric with respect to three different classifiers such as COG, ACOG and ACSSRTGC on two datasets such as Covtype and German. The proposed ACSSRTGC algorithm gives higher sensitivity of 74.25% for Covtype dataset, whereas other methods such as COG and ACOG give only 50.67% and 72.63% respectively.

Fig. 4 shows performance comparison results of specificity metric with respect to three different classifiers such as COG, ACOG and ACSSRTGC on two datasets such as Covtype and German. The proposed ACSSRTGC algorithm gives higher specificity of 68.85% for Covtype dataset, whereas other methods such as COG and ACOG give only 57.015% and 63.67% respectively.

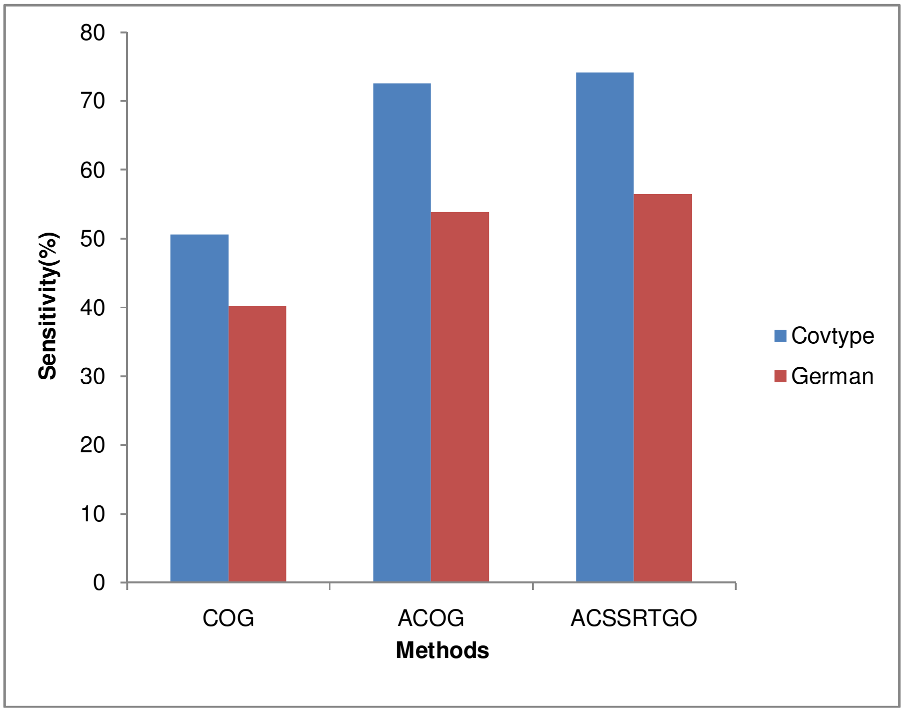

Fig. 3. Sensitivity Results Evaluation of classification methods.

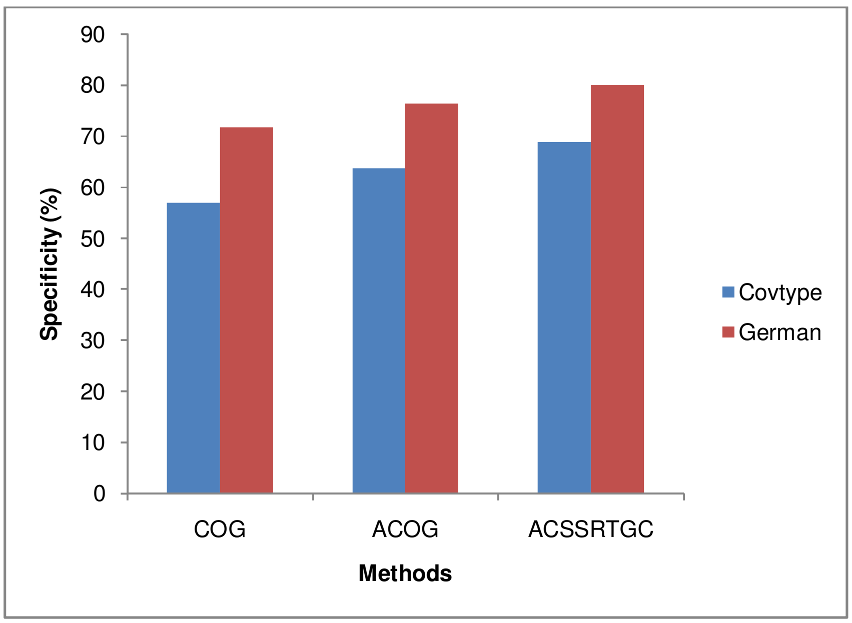

Fig. 4. Specificity Results Evaluation of classification methods.

VI. CONCLUSION AND FUTURE DIRECTION

Motivated by imbalanced dataset issue, propose to introduce Adaptive Cost-Sensitive Sparse Representation Truncated Gradient Online Classification (ACSSRTGC) to empower the cost-critical online classification. The proposed Truncated Gradient can be highly beneficial for resolving an extensive array of imbalanced dataset issue found in online classification. But, the advantages gained of ACSSRTGC is by reducing performance, since it neglects information of correlation among dimensions of sample, which are very significant and inevitable to dataset having great inherent correlation. In addition, apply ACSSRTGC into datasets having two high-dimensions (i.e., Covtype, German), as for tasks of low-dimension, new ACSSRTGC algorithms are adequately quick. In conclusion, all the potential outcomes promise the superiority of the newly introduced algorithms for datasets, where high-dimensional datasets are generally used and are in hugely imbalanced state with respect to class. The works intended for the future directions are: (i) further investigation on deeper concept of cost-critical online learning; (ii) more analysis on sparse computation techniques in problems involving cost-critical online classification.

REFERENCES

[1]. Wu, Q., Wu, H., Zhou, X., Tan, M., Xu, Y., Yan, Y., & Hao, T. (2017). Online transfer learning with multiple homogeneous or heterogeneous sources. *IEEE*

Transactions on Knowledge and Data Engineering, *29*(7), 1494-1507.
[2]. Zhao, P., & Hoi, S. C. (2013). Cost-sensitive online active learning with application to malicious URL detection. In *Proceedings of the 19th ACM SIGKDD international conference on Knowledge discovery and data mining*, 919-927.
[3]. Batista, G. E., Prati, R. C., & Monard, M. C. (2004). A study of the behavior of several methods for balancing machine learning training data. *ACM SIGKDD explorations newsletter*, *6*(1), 20-29.
[4]. Weiss, G. M. (2004). Mining with rarity: a unifying framework. *ACM Sigkdd Explorations Newsletter*, *6*(1), 7-19.
[5]. Japkowicz, N., & Stephen, S. (2002). The class imbalance problem: A systematic study. *Intelligent data analysis*, *6*(5), 429-449.
[6]. Masnadi-Shirazi, H., & Vasconcelos, N. (2010). Risk minimization, probability elicitation, and cost-sensitive SVMs. In *ICML*, 759-766.
[7]. Wang, J., Zhao, P., & Hoi, S. C. (2013). Cost-sensitive online classification. *IEEE Transactions on Knowledge and Data Engineering*, *26*(10), 2425-2438.
[8]. Lee, S., Stephen J.W. (2012). Manifold identification in dual averaging for regularized stochastic online learning. *Journal of Machine Learning Research*. *13*:1705–1744.
[9]. Crammer, K., Dekel, O., Keshet, J., Shalev-Shwartz, S., & Singer, Y. (2006). Online passive-aggressive algorithms. *Journal of Machine Learning Research*, *7*, 551-585.
[10]. Chen, Z., Fang, Z., Fan, W., Edwards, A., & Zhang, K. (2017). CSTG: An Effective Framework for Cost-sensitive Sparse Online Learning. In *Proceedings of the 2017 SIAM International Conference on Data Mining*, 759-767.
[11]. Zhenbing Liu, Chunyang Gao, Huihua Yang, Qijia He, Classification for Class-Imbalance Problem, Scientific Programming, 1 to 9, 2016.
[12]. Ma, Y., & Zheng, T. (2017). Stabilized sparse online learning for sparse data. *The Journal of Machine Learning Research*, *18*(1), 4773-4808.
[13]. Ding, Y., Zhao, P., Hoi, S. C., & Ong, Y. S. (2016). Adaptive Subgradient Methods for Online AUC Maximization, 1- 14.
[14]. Zhang, Q., Zhang, P., Long, G., Ding, W., Zhang, C., & Wu, X. (2016). Online learning from trapezoidal data streams. *IEEE Transactions on Knowledge and Data Engineering*, *28*(10), 2709-2723.
[15]. Yan, Y., Wu, Q., Tan, M., Ng, M. K., Min, H., & Tsang, I. W. (2017). Online heterogeneous transfer by hedge ensemble of offline and online decisions. *IEEE transactions on neural networks and learning systems*, *29*(7), 3252-3263.
[16]. Crammer, K., Kulesza, A., & Dredze, M. (2009). Adaptive regularization of weight vectors. In *Advances in neural information processing systems*, 414-422.
[17]. Zhao, P., Zhang, Y., Wu, M., Hoi, S. C., Tan, M., & Huang, J. (2018). Adaptive cost-sensitive online classification. *IEEE Transactions on Knowledge and Data Engineering*, *31*(2), 214-228.
[18]. Zinkevich, M. (2003). Online convex programming and generalized infinitesimal gradient ascent. In *Proceedings of the 20th international conference on machine learning (icml-03)*, 928-936.
[19]. Luo, H., Agarwal, A., Cesa-Bianchi, N., & Langford, J. (2016). Efficient second order online learning by sketching. In *Advances in Neural Information Processing Systems*, 902-910.
[20]. Liu, Y., Yan, Y., Chen, L., Han, Y., & Yang, Y. (2019). Adaptive Sparse Confidence-Weighted Learning for Online Feature Selection. In *Proceedings of the AAAI Conference on Artificial Intelligence*, 3, 4408-4415.

How to cite this article: Priakanth, P., Thangamani, M. and Gopikrishnan, S. (2020). Adaptive Cost-Sensitive Sparse Representation Truncated Gradient Online Classification (ACSSRTGC). *International Journal on Emerging Technologies*, *11*(2): 283–288.