International Journal of Multidisciplinary Engineering in Current Research

ISSN NO: 2456 - 4265

Volume 5, Issue 2, Febrauary 2020, http://ijmec.com/

Online Weighted Extreme MachineLearning Solution for Active Learning from Unbalanced Data

Dr.J.Chadnra Sekhar[1], Mrs.N.R.L.Prasanna[2], Mr.N.KanthiPriya Darshini[3], Mr.V.K.Pratap[4]

Professor & HOD[1], Assistant Professor[2,3,4]
Department of CSE
NRI INSTITUTE OF TECHNOLOGY, Visadala X Road, Medikonduru(M), Guntur (Dist), Andhra Pradesh 522438.

Abstract

A well-known fact is that active learning may simultaneously increase the quality of a classification model and reduce the complexity of training examples. Several prior research, on the other hand, have shown that an unbalanced distribution of data may readily disturb the performance of active learning. There are several current unbalanced active learning systems that have either poor performance or a large time consumption. Using the extreme learning machine (ELM) classification model, this research presents an effective method dubbed active online-weighted ELM (AOW-ELM). The main contributions of this paper include: 1) a detailed discussion of the reasons why an imbalanced instance distribution can disrupt active learning and its influencing factors; 2) the use of the hierarchical clustering technique to select initially labelled instances in order to avoid the missed cluster effect and the cold start phenomenon as much as possible; and 3) the selection of the weighted ELM (WELM) as the base classifier to ensure the The proposed AOW-ELM approach outperforms many state-of-the-art active learning techniques created expressly for class imbalance in experiments on 32 binary-class data sets with various imbalance ratios.

Index Terms— When it comes to a student's ability to learn and progress, there are a number of factors to consider.

INTRODUCTION

In situations where huge numbers of cases may be cheaply gathered but labelling them is expensive and/or time consuming, ACTIVE learning is a common machine learning methodology [1]. By using active learning, students will be able to learn more effectively. Revisions were made on March 4 and June 27 of this year, and the manuscript was formally accepted on July 3. The National Natural Science Foundation of China, the Natural Science Foundation of Jiangsu Province of China, the China Postdoctoral Science Foundation, the Jiangsu Planned Projects for Postdoctoral Research Funds, and the Qing Lan Project of Jiangsu Province of China all contributed to this work. Digital Object Identifier With iterative human-computer interaction, a classification model may rapidly improve its performance by selecting and classifying just the most important examples. Since human experts are less burdened and training examples are less difficult, active learning's advantages lay in getting a classification model that performs as well as or better than the model with all instances labelled. Active learning models have been studied extensively in the past, and we have a variety of taxonomies for organising these models. It is possible to separate active learning into pool-based [2], stream-based [3] and stream-based models [4]. The former gathers and prepares all unlabeled instances, whilst the latter can only visit a batch of freshly received unlabeled data at a certain time interval.. We have single-mode and batch-mode learning models based on the number of labelled cases in each cycle [5]. The batch-mode model labels all of the unlabeled cases at once, while the single-mode model labels just one unlabeled instance every round, as their names imply. Uncertainty [6], [7], representativeness [8], inconsistency [9], variance [10), and error (11), among others, may be used to rank unlabeled examples. For each significance measure, the criteria for determining which instances are most critical for enhancing the classification model's performance are laid forth. Uncertainty, representativeness, and inconsistency all consider the most important unlabeled instance to be the one closest to the current classification boundary; inconsistency considers the unlabeled instance that has the most predictive divergence among multiple diverse baseline classifiers to be more significant; and uncertainty considers the most important unlabeled instance to be the one closest to the current classification boundary. Active learning models may also be categorised depending on the kind of classifier used in them. In order to meet the needs of active learning, many

International Journal of Multidisciplinary Engineering in Current Research
ISSN NO: 2456 - 4265
Volume 5, Issue 2, Febrauary 2020, http://ijmec.com/

well-known classifiers, including as naive Bayes, k-nearest neighbours, decision trees, multi ple layered perceptrons, logistic regression, support vector machines (SVMs), and extreme learning machines have been created. Additionally, active learning has been used in a wide range of real-world applications, including video annotation, image retrieval, text classification, remote sensing image annotation, and voice recognition, as well as network intrusion detection and bioinformatics, throughout the last decade. With iterative human-computer interaction, a classification model may rapidly improve its performance by selecting and classifying just the most important examples. Since human experts are less burdened and training examples are less difficult, active learning's advantages lay in getting a classification model that performs as well as or better than the model with all instances labelled. Active learning models have been studied extensively in the past, and we have a variety of taxonomies for organising these models. It is possible to separate active learning into pool-based [2], stream-based [3] and stream-based models [4]. The former gathers and prepares all unlabeled instances, whilst the latter can only visit a batch of freshly received unlabeled data at a certain time interval.. We have single-mode and batch-mode learning models based on the number of labelled cases in each cycle [5]. The batch-mode model labels all of the unlabeled cases at once, while the single-mode model labels just one unlabeled instance every round, as their names imply. Uncertainty [6], [7], representativeness [8], inconsistency [9], variance [10), and error (11), among others, may be used to rank unlabeled examples. For each significance measure, the criteria for determining which instances are most critical for enhancing the classification model's performance are laid forth. Uncertainty, representativeness, and inconsistency all consider the most important unlabeled instance to be the one closest to the current classification boundary; inconsistency considers the unlabeled instance that has the most predictive divergence among multiple diverse baseline classifiers to be more significant; and uncertainty considers the most important unlabeled instance to be the one closest to the current classification boundary. Active learning models may also be categorised depending on the kind of classifier used in them. In order to meet the needs of active learning, many well-known classifiers, including as naive Bayes, k-nearest neighbours, decision trees, multi ple layered perceptrons, logistic regression, support vector machines (SVMs), and extreme learning machines have been created. Additionally, active learning has been used in a wide range of real-world applications, including video annotation, image retrieval, text classification, remote sensing image annotation, and voice recognition, as well as network intrusion detection and bioinformatics, throughout the last decade. Recent studies show that active learning fails when it's applied to data with a skewed distribution of class members. Active learning, like conventional supervised learning, does not shy away from the issue of an unequal number of students in each class. A number of prior research have attempted to solve this issue using a variety of approaches. First to detect this issue, the authors Zhu and Hovy [25] used numerous sampling approaches into the active learning process to regulate the distribution of labelled cases between the minority and majority classes. RUS (Random Undersampling), ROS (Random Oversampling), and BOOTOS (bootstrap-based Oversampling) all appeared in the study. Both ROS and BootOS may improve learning performance, but RUS is often poorer than the original active learning algorithm, according to the authors, who also found that RUS is more prone to overfitting. As another prominent class imbalance learning strategy, Bloodgood and Vijay-Shanker [26] used cost-sensitive learning in order to address skewed data distribution during active learning. In specifically, the basic learner was cost-sensitive SVM (CS-SVM), empirical costs were allocated based on the prior imbalance ratio, and two classical stopping criteria, i.e., the minimal error and the highest confidence, were used to discover the suitable stopping condition for active learning. In spite of the time-consuming nature of training an SVM and the lack of online learning, the approach is very successful. For instance, Tomanek [31] and Hahn [32] proposed two methods based on the inconsistency significance measure: balanced-batch

International Journal of Multidisciplinary Engineering in Current Research
ISSN NO: 2456 - 4265
Volume 5, Issue 2, Febrauary 2020, http://ijmec.com/

active learning (AL-BAB) and AL-BOOD, where the former selects from 5n new labels n labelled instances that are class balanced on each round of active learning, while the latter modifies the equation of voting entropy to focus on the minority class. When compared to RUS, AL-BAB has a lot in common, but it may be much worse and use up even more labelled resources since it relies on so many different base learners (semantic learning) to compute the voting entropy of predictive labels, which necessarily increases the computing burden. Consequently, in Section V, we did not compare our suggested technique to the methods listed above. Additionally, research has been done on how to address the issue of class imbalance by using active learning. Ertekin et al. [32, 33] found that the imbalance ratio is substantially lower at the border between two distinct classes, therefore active learning may effectively mitigate the detrimental consequences of unbalanced data distribution. Of put it another way, they see active learning as a particular approach to sampling. It is also advised that since they used SVM as their base learner, they should utilise an early stopping criteria of margin depletion to verify the stopping condition. As a summary of current active learning algorithms, we observed that they either suffer from poor classification performance or high time consumption difficulties in the case of uneven data distributions. This paper's goal is to provide a method that is both effective and efficient. AOW-ELM is the name of the suggested method, and it should be used in a batch-mode active learning scenario with an uncertainty significance measure and an ELM classifier. Because of its superior generality and classification performance compared to SVM and MLP [34], [35], as well as its potential to significantly reduce training time compared to other classifiers [36], we choose ELM as the baseline classifier for active learning. The weighted ELM (WELM) [37] is used as the basis learner in AOW-ELM to overcome the class imbalance issue in active learning. A framework for active learning is then built using our earlier paper's AL-ELM algorithm [21]. The next step is to develop a weight update algorithm and an efficient online learning mode for WELM. Finally, we provide a more flexible and effective early stop ping criteria that benefits from the notion of the margin exhaustion requirement. To that end, we look at the impact of three key distribution characteristics, namely the class imbalance ratio, class overlapping, and tiny disjunction, to see how they affect active learning. Furthermore, in order to prevent both the cold start and missing cluster impact, we recommend using the clustering approaches to pick a seed set that is already labelled. Using 32 binary-class unbalanced datasets, the findings show that the suggested algorithmic framework is typically more effective and efficient than various active learning algorithms that were explicitly created for the class imbalance situation. Listed below are the sections of this document. Prior to reading this article, you should have a basic understanding of the subject matter. Using synthetic data sets with varied distributions, we examine why active learning might be ruined by skewed instance distributions in Section III. In Section IV, we go into the specifics of the algorithmic framework we've come up with. Section V summarises and analyses the findings of the experiment. Finally, Section VI sums up the paper's findings and points to further research.

PRELIMINARIES

Preliminaries are shown here, covering the fundamental flow route of pool-based active learning, ELM and WESM as well as online sequential ELM and ELM-based active learning (AL-ELM). It is in Section IVA of this work where the main algorithmic paradigm is described. Pool-Based Active Learning's Flow Path The pool-based scenario is more typical in real-world applications, as described in Section I, and may be separated into two types according to distinct means of inputting the unlabeled data: pool-based [2], [3] and stream-based [4]. Unlabeled cases in a pool are pre-prepared, and then a random subset is selected.

ISSN NO: 2456 - 4265
Volume 5, Issue 2, Febrauary 2020, http://ijmec.com/

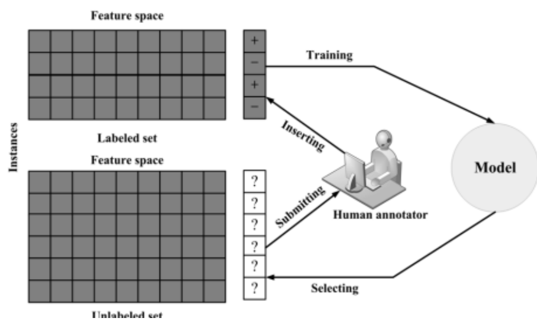


Fig. 1. Flowchart of the pool-based active learning

Active learning repeatedly retrieved and tagged by human specialists. Furthermore, the classification model is able to continually choose the most relevant examples from a pool of unlabeled examples. The pool-based active learning scenario is the subject of this research. Pool-based active learning is shown in Fig. 1 as a simple flow route. In this diagram, we can see that the flow route is divided into four sections and is a complete loop. It is necessary to train a classification model on the labelled data each time. A subset of the unlabeled instances is then selected for human annotation by the model, which ranks them according to their importance and sends only the most important ones to the annotators for further labelling. The method outlined above is called "active learning," and it is repeated until a certain condition is met. How to estimate, rank, and extract important unlabeled instances using the classification model is clearly the critical step in the flow route described above, as it directly affects the quality of active learning. B. The Extreme Learning Machine. Electronic Learning Mechanism (ELM) was developed by Huang et al. [34, 35] to train single-hidden layer feed forward neural networks (SLFNs). It is the random creation of concealed nodes that makes ELM stand out from SLFN's typical learning methods. There are no iterative adjustments required by ELM in order to get optimum values, therefore it learns quicker and generalises better. Compared to SVM and MLP, previous studies have shown that ELM can yield superior generality and classification performance or at least be equivalent [34]–[36]. However, ELM requires tenths or hundredths of the training time. The ith training instance may be written as (xi, ti), where xi is a n1 input vector and ti is the associated m1 output vector in a classification problem with N training cases. All weights and biases on the L hidden nodes in ELM are created at random, and we may assume that this is the case. h(xi) = [h1(xi), h2(xi),..., hL (xi)] represents the hidden layer output for the instance xi. activation function mapping (the most popular sigmoid function is used throughout this paper). ELM's mathematical model might be referred to as

$$\beta = H^{\dagger}T = \begin{cases} H^T(HH^T)^{-1}T, & \text{when } N \le L \\ (HH^T)^{-1}H^TT, & \text{when } N > L \end{cases}$$

C is the penalty factor that represents the tradeoff between the reduction of training mistakes and the maximum of generalisation ability, which is denoted by I = [i,1], [i,2,...,..., i,m].] The Karush–Kuhn–Tucker theorem [38] may be used to solve a common quadratic programming issue. For (3), the answer may be summarised as follows.

The D. Machine for Online Sequential Extreme Training When Liang et al. introduced OS-ELM in 2006, they were referring to an online sequential learning mode of ELM. For training with sequentially obtained data, OS ELM uses extended recursive least squares, which may either be received one at a time or chunk by chunk. The update rule for the output layer weight matrix,, may be expressed in terms of the following:

Where Hk+1 and Tk+1 represent new observations in the (k + 1)th chunks, and (k) and (k+1) designate the output layer weight matrix after receiving the kth and (k + 1)th chunks, respectively, in this case. The formula for Pk+1 may be found as follows: where When fresh observations are made in (k+1)th chunks, Hk+1, Tk+1 and (k) are the hidden layer output matrix and the target matrix, respectively. The output layer weight matrix after receiving kth and (k+1)th pieces is referred to as (k). The formula for Pk+1 may be found as follows: ELM outputs and Bayes

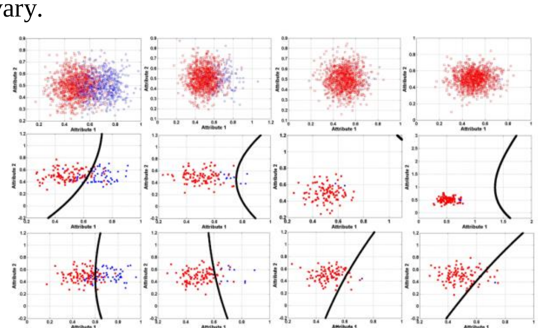
International Journal of Multidisciplinary Engineering in Current Research
ISSN NO: 2456 - 4265
Volume 5, Issue 2, Febrauary 2020, http://ijmec.com/

classifier posterior probabilities are linked via a mapping connection. The posterior probabilities corresponding to the distinct classes are more unclear and important for modelling the classifier when the actual outputs on different output nodes are more approximate. Classification model quality may be improved by labelling the most doubtful cases, such as those closest to the present classification hyperplane, as discussed in Section I. The sigmoid function is used to transform ELM's nonprobabilistic outputs into probabilistic outputs as follows:

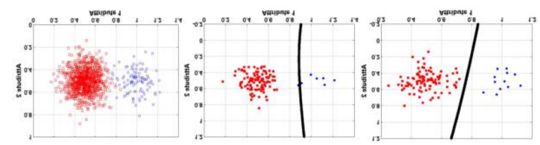
$$P(t_i | f_i(x)) = \frac{1}{1 + \exp(-f_i(x))}, \quad i = 1, 2, \ldots, m$$

INVESTIGATION INTO THE DISTRIBUTION

A skewed distribution of data may quickly devastate typical active learning systems. To understand why, we look at data distribution. Class imbalance, class overlapping, and minor disjunction are all aspects that we examine while creating six synthetic 2-D binary-class unbalanced data sets. Table I summarises the distributions of these six data sets. Table I shows that data sets D1–D4 have the same distribution, although class imbalance ratios ranging from 2:1 to 200:1 may be seen. However, despite the fact that the class imbalance ratio in data set D5 is identical to that in data set D2, their distributions vary.



Data sets one through four are shown in Fig. 2 in their original distribution. Classification hyperplane generated by ELM on 10% randomly extracted initial labelled examples (L = 100, C = 210) is shown in the second row. A WELM classification hyperplane (L=100, C=210) on 10% of the original labelled examples is shown in the third row. One data set is represented by each column (D1 to D4), whereas and signify the majority class's unlabeled and labelled instances while and denote the minority class's unlabeled and labelled instances (from left to right). (See Fig. 2) The first four data sets are shown in their original distribution in the first row. ELM (L = 100, C = 210) generated a classification hyperplane on 10% of the original labelled examples. A WELM classification hyperplane (L=100, C=210) on 10% of the original labelled examples is shown in the third row. the unlabeled and the labelled instances of the majority class are shown in and the unlabeled and labelled examples of the minority class are shown in accordingly, each column denoting a data set (D1–D4) (from left to right). Data sets one through four are shown in Fig. 2 in their original distribution. Classification hyperplane generated by ELM on 10% randomly extracted initial labelled examples (second row) (L = 100, C = 210) Classification hyperplane obtained by WELM on 10% randomly extracted initial labelled examples (L = 100, C = 210) is shown in row 3. Each row represents a different set of data (D1–D4). as well as designate the majority class's and labelled instances, respectively; the same is true for and of the minority class (from left to right).



ELM (L = 100, C = 210) and WELM (L = 100, C = 210) generated the original distribution and classification hyperplane on 10% randomly extracted initial labelled cases of the D5 data set, respectively. In the majority class, and signify the unlabeled and labelled examples, while and denote the unlabeled and labelled instances of the minority class,

ELM, for example, tends to yield a hyperplane closer to the majority class on the big margin class imbalance data set than on the short margin data set,

International Journal of Multidisciplinary Engineering in Current Research
ISSN NO: 2456 - 4265
Volume 5, Issue 2, Febrauary 2020, http://ijmec.com/

as seen in Fig. 2 and 3. To put it another way, as the margin grows, the damage will diminish. To counter this damage to regular classifiers, a balancing control approach must be implemented (compare the second subgraph with the third subgraph in Fig. 3). Within-class subclusters, also known as minor disjunctions, are examined as a more sophisticated data distribution component. Subclasses and subdistributions in a single class might have varying numbers of subclasses and subdistributions. We confront both the issue of "between-class imbalance" and the problem of "within-class imbalance" on the skewed data set with minor disjunctions. If we take a look at the D6 data set, there are three clusters within each class with an imbalance ratio of 6:3:1. It is conceivable to permanently misclassify all instances in the tiny subclusters if none of the small subclusters are included in the first labelled set (seed set). The missing cluster effect [45] is the name given to this issue as a result of human error. On the large margin class imbalance data set, regular classifiers, such as ELM, generally produce a hyperplane that is closer to the majority class than that produced on the smaller margin data set (Fig. 2 and Fig. 3). We further observe that on the large margin data set, the hierarchical clustering technique is used to subtly explore the distribution structure of collected unlabeled instances. In other words, the harmfulness will diminish as the margin size increases. Even so, regular classifiers may still be harmed by it, hence a balancing management method must still be used (compare the second subgraph with the third subgraph in Fig. 3). Within-class subclusters, also known as minor disjunctions, are examined as a more sophisticated data distribution component. Subclasses and subdistributions in a single class might have varying numbers of subclasses and subdistributions. We confront both the issue of "between-class imbalance" and the problem of "within-class imbalance" on the skewed data set with minor disjunctions. If we take a look at the D6 data set, there are three clusters within each class with an imbalance ratio of 6:3:1. It is conceivable to permanently misclassify all instances in the tiny subclusters if none of the small subclusters are included in the first labelled set (seed set). The missing cluster effect [45] is the name given to this issue as a result of human error. The hierarchical clustering technique is used to investigate the distribution structure of the collected unlabeled instances, and by selecting the column in Fig. 2 with Fig. 3, we see that the regular classifier, e.g., ELM, generally produces a hyperplane that is closer to the majority class than the hyperplane produced on the small margin data set. In other words, the harmfulness will diminish as the margin size increases. Even so, regular classifiers may still be harmed by it, hence a balancing management method must still be used (compare the second subgraph with the third subgraph in Fig. 3). Within-class subclusters, also known as minor disjunctions, are examined as a more sophisticated data distribution component. Subclasses and subdistributions in a single class might have varying numbers of subclasses and subdistributions. We confront both the issue of "between-class imbalance" and the problem of "within-class imbalance" on the skewed data set with minor disjunctions. If we take a look at the D6 data set, there are three clusters within each class with an imbalance ratio of 6:3:1. It is conceivable to permanently misclassify all instances in the tiny subclusters if none of the small subclusters are included in the first labelled set (seed set). The missing cluster effect [45] is the name given to this issue as a result of human error. The hierarchical clustering approach is used to investigate the distribution structure of the gathered unlabeled examples and to make more exact selections to solve this issue.

AOW-ELM ALGORITHM MODEL

An Extreme Learning Machine that may be used online. If you've been paying attention, you've seen that active learning uses an iterative process, in which new labelled instances are added to the set each time iteration is completed. Retraining the categorization model after each round would take a considerable amount of time. To implement active learning, an online learning algorithm must be used. The incremental class imbalance learning issue was addressed by Mirza et al. [47] using an online sequential WELM method. Even though ELM may

International Journal of Multidisciplinary Engineering in Current Research
ISSN NO: 2456 - 4265
Volume 5, Issue 2, Febrauary 2020, http://ijmec.com/

be generalised, this method only uses the least squares version of ELM and not its optimization version, limiting its generalizability. Its repetitive weight adjusting and ranking method, on the other hand, adds to the waste of time. An online imbalance learning technique based on a kernel was given in [48]. The method is helpful in combating online learning that is unbalanced and prone to drifting conceptualizations. Due to the fact that its instance weight is correlated with the number of training instances belonging to each class that have been acquired thus far, its applicability in our active learning scenario is limited. As a result, new instances will gradually lose weight, while the old ones will be highlighted. It's clear that it doesn't meet our needs. This work proposes a new online sequential WELM method based on the optimization version to eliminate these issues. Like the OS-ELM algorithm [39], it was derived in a similar manner. In light of (8), we may say

$$\beta = \left(\frac{I}{C} + HWH^T\right)^{-1} H^T WT$$
$$= \left(\frac{I}{C} + (\sqrt{WH})^T(\sqrt{WH})\right)^{-1} (\sqrt{WH})^T(\sqrt{WT}).$$

Then, (14) can be rewritten as

$$\beta = \left(\frac{I}{C} + U^T U\right)^{-1} U^T V$$

where

$$U = \sqrt{WH}$$
$$V = \sqrt{WT}.$$

VI. CONCLUSION

AOW-ELM is a solution to the issue of active learning in a class imbalance situation that we offer in this work. Several variables contribute to the harmfulness of skewed data distribution, which we discover is really a mixture of these aspects. Using hierarchical clustering, the possible missing cluster effect and cold start phenomena may be efficiently addressed by extracting early representative instances into a seed set in advance. Using existing benchmark methods, the suggested AOW-ELM algorithm outperforms them all in addressing the issue of active learning in a class imbalance situation. The following are some of the advantages of the AOW-ELM algorithm. A weight update rule that is resilient is one of its advantages. 2) It runs quickly and linearly as the number of training examples increases. The early halting criteria is adjustable. A wide range of data sets may be used with it. Active learning on multiclass unbalanced data sets will be a major focus of our future work. We'll also look at active learning solutions for dealing with unlabeled and unbalanced data streams as well as managing idea drifts.

REFERENCES

[1] B. Settles, “Active learning literature survey,” Dept. Comput. Sci., Univ. Wisconsin-Madison, Madison, WI, USA, Tech. Rep. 1648, Jan. 2010.

[2] M. Sugiyama and S. Nakajima, “Pool-based active learning in approximate linear regression,” Mach. Learn., vol. 75, no. 3, pp. 249–274, 2009.

[3] A. K. McCallumzy and K. Nigam, “Employing EM and poolbased active learning for text classification,” in Proc. ICML, 1998, pp. 359–367.

[4] J. Smailovi´c, M. Grˇcar, N. Lavraˇc, and M. Žnidarši˘c, “Stream-based active learning for sentiment analysis in the financial domain,” Inf. Sci., vol. 285, pp. 181–203, Nov. 2014.

[5] S. C. H. Hoi, R. Jin, J. Zhu, and M. R. Lyu, “Batch mode active learning and its application to medical image classification,” in Proc. ICML, 2006, pp. 417–424.

[6] D. Lewis and J. Catlett, “Heterogeneous uncertainty sampling for supervised learning,” in Proc. ICML, 1994, pp. 148–156.

[7] M. Lindenbaum, S. Markovitch, and D. Rusakov, “Selective sampling for nearest neighbor classifiers,” Mach. Learn., vol. 54, no. 2, pp. 125–152, 2004.

[8] S. Dasgupta and D. Hsu, “Hierarchical sampling for active learning,” in Proc. ICML, 2008, pp. 208–215.

[9] R. Wang, S. Kwong, and D. Chen, “Inconsistency-based active learning for support vector machines,” Pattern Recognit., vol. 45, no. 10, pp. 3751–3767, 2012.

[10] S. C. H. Hoi, R. Jin, and M. R. Lyu, “Large-scale text categorization by batch mode active learning,” in Proc. WWW, 2006, pp. 633–642.

International Journal of Multidisciplinary Engineering in Current Research

ISSN NO: 2456 - 4265

Volume 5, Issue 2, Febrauary 2020, http://ijmec.com/

[11] N. Roy and A. McCallum, "Toward optimal active learning through Monte Carlo estimation of error reduction," in Proc. ICML, 2001, pp. 441–448.

[12] D. D. Lewis and W. A. Gale, "A sequential algorithm for training text classifiers," in Proc. ACM SIGIR, 1994, pp. 3–12.

[13] J. He and J. G. Carbonell, "Nearest-neighbor-based active learning for rare category detection," in Proc. NIPS, 2008, p. 633–640.

[14] A. Krogh and J. Vedelsby, "Neural network ensembles, cross validation, and active learning," in Proc. NIPS, 1995, pp. 231–238.

[15] K. Fukumizu, "Statistical active learning in multilayer perceptrons," IEEE Trans. Neural Netw., vol. 11, no. 1, pp. 17–26, Jan. 2000.

[16] A. I. Schein and L. H. Ungar, "Active learning for logistic regression: An evaluation," Mach. Learn., vol. 68, no. 3, pp. 235–265, 2007.

[17] S. Tong and D. Koller, "Support vector machine active learning with applications to text classification," J. Mach. Learn. Res., vol. 2, pp. 45–66, Nov. 2001.

[18] S. Tong and E. Chang, "Support vector machine active learning for image retrieval," in Proc. 9th ACM Int. Conf. Multimedia, 2001, pp. 107–118.

[19] Y. Qi and G. Zhang, "Strategy of active learning support vector machine for image retrieval," IET Comput. Vis., vol. 10, no. 1, pp. 87–94, 2016. [20] C. Pan, D. S. Park, H. Lu, and X. Wu, "Color image segmentation by fixation-based active learning with ELM," Soft Comput., vol. 16, no. 9, pp. 1569–1584, 2012.

[21] H. Yu, C. Sun, W. Yang, X. Yang, and X. Zuo, "AL-ELM: One uncertainty-based active learning algorithm using extreme learning machine," Neurocomputing, vol. 166, pp. 140–150, Oct. 2015. [22] H. Liao, L. Chen, Y. Song, and H. Ming, "Visualization-based active learning for video annotation," IEEE Trans. Multimedia, vol. 18, no. 11, pp. 2196–2205, Nov.