Deep Learning with MCA-based Instance Selection and Bootstrapping for Imbalanced Data Classification

Sheng Guan[1], Min Chen[2], Hsin-Yu Ha[1], Shu-Ching Chen[1], Mei-Ling Shyu[3], and Chengde Zhang[4]
[1]School of Computing and Information Sciences
Florida International University, Miami, FL 33199, USA
[2]Computing and Software Systems, School of STEM
University of Washington Bothell, Bothell, WA 98011, USA
[3]Department of Electrical and Computer Engineering
University of Miami, Coral Gables, FL 33146, USA
[4]School of Information and Safety Engineering
Zhongnan University of Economics and Law, Hubei 430073, China

sguan005@cs.fiu.edu, minchen2@uw.edu, hha001@cs.fiu.edu, chens@cs.fiu.edu, shyu@miami.edu, chengde66@gmail.com

Abstract— **In this paper, we propose an extended deep learning approach that incorporates instance selection and bootstrapping techniques for imbalanced data classification. In supervised learning, classification performance often deteriorates when the training set is imbalanced where at least one of the classes has a substantially fewer number of instances than the others. We propose to use adaptive synthetic sampling approach (ADASYN) to generate synthetic instances for the minority class. A data pruning process based on multiple correspondence analysis (MCA) is then performed to identify a subset of synthetic instances that are most suitable to supplement the existing minority instances. This results in a relatively more balanced training dataset which is then bootstrapped and fed into the convolutional neural networks (CNNs) for classification. Furthermore, we propose to use low-level features preprocessed by principal component analysis (PCA), instead of the commonly used raw signal data, as the input to CNNs to reduce the computational time. The experimental results show the effectiveness of our framework in classifying 54 TRECVID concepts with different imbalanced levels by comparing with other state-of-the-art methods.**

Keywords-Classification, imbalanced data, bootstrapping, Convolutional Neural Network (CNN), supervised learning, Multiple Correspondence Analysis (MCA)

I. INTRODUCTION

In many applications [1]-[10], large amounts of data are generated with a skewed distribution (or called an imbalanced dataset) where at least one of the classes is represented by a significantly fewer number of instances than the others. In addition, the rare instances that constitute the minority class are generally considered as the concept of interest. For instance, in biomedical research, the data instances for different kinds of malignant cancers are generally very rare compared to normal cells. However, these rare instances require special attentions and it is essential to predict their presence or classify them as accurate as possible [11]. Consequently, the ratio of the minority instances to the majority instances is often called P/N ratio (i.e., positive to negative ratio) to indicate the degree of imbalance in the dataset.

Most mainstream classifiers are modeled based on the statistics of the training data, assuming that the class distribution is balanced or misclassification costs are equal [12]. Therefore, they often perform poorly in imbalanced data classification, where the models are biased towards the majority class (negative class) with its data instances overshadowing those in the minority class (positive class). It is even more challenging when the dataset is multimedia data due to its diverse media types and spatio-temporal characteristics [13]-[19].

Recently, this problem of imbalanced data classification has attracted significant research efforts in machine learning, artificial intelligence, data mining and related areas [20]-[23]. For instance, data manipulation methods are proposed in [24] to change the distribution of the training set to improve the classification performance on imbalanced datasets. The common data manipulation methods can be grouped into two categories: over-sampling or under-sampling methods. The over-sampling methods tend to duplicate existing positive instances or generate synthetic ones to expand the positive instance pool but may result in overfitting. The under-sampling methods select a part of negative samples to reduce the imbalanced degree of the training set but may lead to the loss of information.

On the other hand, the selection of classifiers also plays an important role to improve the classification performance on imbalanced data. Deep learning approaches [25] such as convolutional neural networks (CNNs), inspired by the research in neuroscience that human brains perform well in tasks like object recognition, are able to extract more abstract and high-level features from the data and are believed to excel many traditional classifiers. However, their performance can actually be rather poor in imbalanced data classification as we have observed in our empirical study and they

are often too computationally demanding to apply on large multimedia datasets.

In this paper, an extended CNN-based deep learning framework is proposed to improve imbalanced multimedia data classification. It consists of three components. First, the adaptive synthetic sampling approach (ADASYN) is adopted to generate synthetic instances for the minority class. As discussed in [26], ADASYN is motivated by several state-of-the-art synthetic sampling methods to change the initially imbalanced data distribution and it excels in reducing the learning bias. However, different combinations of these synthetic instances can lead to very diverse classification performances. To our best knowledge, no mechanism has been proposed to enhance ADASYN with the capability of selecting suitable synthetic instances for better results. Therefore, in the second component, a novel MCA (multiple correspondence analysis)-based supervised approach is proposed to improve the synthetic instance pool to fit the unique properties of the data. The selected synthetic instances are then used as additional minority instances to balance the training dataset, which is then passed to the third component to be bootstrapped and fed into the convolutional neural networks (CNNs) for classification. Here, the bootstrapping method aims to further adjust the distribution of the instances to improve the CNNs performance. In addition, to address the issue that deep learning approaches such as CNNs are usually computationally expensive in processing raw data instances, we propose to extract low-level features, preprocess them using principal component analysis (PCA) to reduce the feature dimension, and feed into CNNs to speed up the training process.

The rest of this paper is organized as follows. In Section II, related work in imbalanced data classification and deep learning is discussed. Section III introduces the proposed framework and its components in details. Experimental results and analyses are presented in Section IV. Finally, Section V concludes this paper.

II. RELATED WORK

A. Adaptive Synthetic Sampling Approach

There are several types of data manipulation techniques to counter the effect of imbalanced datasets [27]-[28]. Among them, ADASYN has been shown to be effective in reducing the bias in the imbalanced dataset [26].The key idea of ADASYN is to use a density distribution as a criterion to decide the number of synthetic instances that need to be generated for each minority data instance. It can adaptively shift the classification decision boundary toward the synthetic data to compensate for the skewed distributions. A detailed algorithm description is in [26].

The simulation analysis conducted in [26] proves that ADASYN can outperform many other methods including SMOTE, a classical synthetic sampling method, decision tree algorithm, and others, for all test benchmarks. Hence, we choose to integrate and extend ADASYN in our framework to handle imbalanced learning problems.

B. Convolutional neural network

As a well-known deep learning method, convolutional neural networks (CNNs) were first proposed by Yann LeCun and Yoshua Bengio [29] who embraced the idea of using various types of neurons organized within one network as an artificial intelligence approach to recognize and process visual patterns. In CNNs, each neuron has its specific functions in image processing so that CNNs are capable of processing image data with the minimum or no preprocessing. CNNs combine three architectural ideas to ensure some degree of shift and distortion invariance: local receptive fields, shared weights, and spatial subsampling [29]. With local receptive fields, neurons can extract elementary image features such as oriented edges, corners, etc. and combine them as the input for the higher layers to detect more complex features. Taking into account that the statistics of one part of a natural image are similar to other parts, elementary feature detectors that are useful on one part are likely to be applicable to the entire image. Therefore, it is reasonable to set a group of units as the receptive fields with identical weight vectors in each neuron to form a small size kernel. The outputs of each neuron constitute a feature map. A convolutional layer is composed of several feature maps (corresponding to several neurons with shared weight vectors), so that multiple features can be extracted in each convolutional layer. The principle of shared weights in CNNs significantly reduces the number of free parameters in model training and improves its generalization ability [30]. To further reduce the computation task, every neuron of the convolutional layer is connected only to a small subset of the lower layer instead of the whole layer in CNNs. Once a feature is detected, its exact location becomes less important as long as its position relative to other features is preserved. Therefore, a convolutional layer can be followed by a spatial subsampling layer to compute its aggregated statistics to reduce the sensitivity of the output to shifts and distortions. After several convolutional and spatial subsampling layers, the high-level reasoning in the neural network is done via fully connected layers [31]. Each fully connected layer computes the dot product of its input and weights, adds a bias, and applies a squashing function as a classifier to the lower layer outputs.

Recently, CNNs have been used in many fields, including speech recognition, vehicle detection, emotion recognition, human action recognition, and traffic sign recognition [32]-[35]. Encouraged by these results, we propose to extend CNNs for imbalanced multimedia data classification.

III. FRAMEWORK

Deep learning has shown to achieve huge success in many research areas. However, very few of them attempted to improve the performance of imbalanced multimedia data classification. In fact, as will be shown in Section IV, the performance is usually unsatisfactory when deep learning is applied directly to a skewed dataset. The reason is that most deep learning approaches, including CNNs, split the training dataset into several mini-batches during training (see detailed descriptions in Section III.C). It is expected that some

of these mini-batches may have no positive instance (called "positive" as it is of users' interest) from the minority class in an imbalanced dataset, which brings the bias towards the negative instances (from the majority class) for the training model. To address this issue, our proposed framework consists of three components: synthetic sampling, instance selection, and deep learning with bootstrapping.

A. ADASYN

The adaptive synthetic sampling approach (ADASYN) is adopted and extended in our framework based on two considerations. First, ADASYN has been proven as an effective method to tackle the imbalanced dataset problem. Second, the synthetic instances generated in ADASYN can be used to increase the total number of positive instances available in the training dataset, which reduces the chance that the same set of positive instances are repeatedly added to multiple mini-batches for CNN training to avoid overfitting.

In brief, ADASYN can generate synthetic positive instances based on the analysis of the whole training dataset with the label information. After m_s synthetic positive instances are generated based on this supervised algorithm, they can be combined with original positive instances m_o to generate a bigger positive instance pool m_f for CNN training, where $m_f = m_o + m_s$.

However, one limitation of applying ADASYN in our framework is that, to our best knowledge, no rules have been defined to determine the value of m_s (i.e., how many synthetic positive instances should be generated and used for each concept). In fact, as a powerful synthetic sampling method, ADASYN can generate as many positive instances as needed (up to $n - m_o$, where n is the number of negative instances in the training dataset) which may or may not lead to an optimal results in CNN classification. Therefore, we propose to generate $(n - m_o) * \beta$ synthetic positive instances in ADASYN. Here, β is a real number which is less than 1. In Section IV, β is set to be 0.01, considering $n >> m_o$ in our training dataset. Then a novel multiple correspondence analysis (MCA) is applied to analyze these synthetic positive instances and identify the most suitable m_s ones for performance improvement as will be discussed in the next section.

B. Integrating MCA with ADASYN

The idea is to use MCA as a pruning tool to assign each instance a score to reflect its relevance to the majority class or the minority class in the training dataset by utilizing the label information. Specifically, MAC will be applied to a dataset that consists of all the negative instances from the original training dataset and all the synthetic positive instances generated from ADASYN. Each instance may be represented by a vector of numerical values (e.g., raw data values, low-level features, etc.). In our framework, low-level features are used to improve the framework efficiency as discussed earlier (with more details in Section III.D). To apply MCA, each feature values are first discretized into several partitions (i.e., feature-value pairs). An example is shown in Table I, assuming there are F features in total.

TABLE I. EXAMPLES FOR NOMINAL DATA INSTANCES

$\mathbf{Feature_1}$ $(\mathbf{A_1})$	$\mathbf{Feature_2}$ $(\mathbf{A_2})$	… …	$\mathbf{Feature_F}$ $(\mathbf{A_F})$	**Class**
A_1^2	A_2^3	…	A_F^2	C_p
A_1^1	A_2^1	…	A_F^3	C_n
…	…	…	…	…
A_1^3	A_2^4	…	A_F^1	C_n
…	…	…	…	…

As can be seen from Table I, each instance occupies one row in the table and is represented by a set of feature-value pairs A_i^j (i.e., j^{th} partition in the i^{th} feature) with the class label at the last column (C_p or C_n in two-class classification, where $C_p = 1$ representing positive class and $C_n = 0$ for negative class). MCA is then used to capture the correlation among more than two variables in Table I. It projects the feature value space into the principle component space and calculates the cosine of the inner product angle ($angle_i^j \in [0, 180]$) between each feature-value pair (A_i^j) and each class of the training dataset to represent their level of correlation. In other words, A_i^j has a higher (or lower) correlation relationship with the positive class if $angle_i^j$ for (A_i^j, C_p) is smaller (or bigger) than 90-degree, respectively. If $angle_i^j$ is equal to 90-degree, then A_i^j is equally correlated with both positive and negative classes. Accordingly, the weight of each feature-value pair A_i^j is computed via Equation (1).

$$WEIGHT_i^j = (180 - angle_i^j)/90. \quad (1)$$

The final score $SCORE_k$ for each row k (i.e., k^{th} training instance) is then calculated as the summation of all its feature-value pair weights as shown in Equation (2), where j is the corresponding partition for each feature.

$$SCORE_k = \sum_{i=0}^{F} WEIGHT_i^j; \quad (2)$$

We rank the synthetic positive instances in a descending order based on their score values and choose top m_s instances for each concept to supplement the m_o positive instances in the original training dataset. Here, we adopt the idea presented in [28] to set $m_s = k * m_o$ (k is normally set to be 1 or 2) with the restriction of $m_f < n$ where $m_f = m_o + m_s$ and n is the total number of the negative instances. Note that if a dataset is severely imbalanced, i.e., $m_o << n$, we will get an m_f that is still far smaller than n so the dataset remains imbalanced even after the ADASYN and MCA steps.

C. CNNs with bootstrapping

To further improve the classification performance on imbalanced datasets, we propose to extend the CNN framework with a bootstrapping method. The bootstrapping method is formally defined as follows.

Let n and m_f be the numbers of negative and positive instances, respectively, in the revised training dataset after the ADASYN and MCA steps. They will then be used to generate a set of mini-batches, which each contains totally S instances with S_m positive and S_n negative instances, for CNNs. The main idea of our algorithm is illustrated in Table II. In step 1, n negative instances are divided into N distinct subsets $X = \{x_i | \, i = 1, 2, \ldots N\}$ where $x_i \cap x_j = \emptyset$ when $i \neq j$, $size(x_i) = S_n$, and $N = \lfloor n/S_n \rfloor$. In other words, any negative instance can be contained in at most one of these subsets. Then, in steps 7-11, all m_f positive instances are divided into M distinct groups ($M = \lceil m_f/S_m \rceil$). If S_m is not a divisor of m_f, we have (M-1) groups with each group having S_m positive instances and one group containing ($m_f - S_m * (M-1)$) positive instances. Each of the groups (with S_m positive instances) will combine with S_n different negative instances (i.e., each negative subset x_i is used only once) to form a mini-batch in step 12. We will randomly regenerate another M groups (steps 5-6) and repeat the same process until N mini-batches are generated.

Using the random partition and selection method, this process ensures that each positive instance has a relatively equal opportunity to be selected and reduces the chance of having the exact same set of positive instances in multiple mini-batches to avoid overfitting.

TABLE II. OVERALL PROCESS OF THE EXTENDED CNN

PSEUDO CODE OF CNN WITH BOOTSTRAPPING
Input: negative set *NG* containing n negative instances, positive set *PS* with m_f positive instances
Output: N mini-batches $MB = \{mb_i | \, i = 1, 2, \ldots N\}$ for CNNs
1. Divide n negative instances into N subsets $X = \{x_i | \, i = 1, 2, \ldots N\}$, each with S_n negative instances
2. Set *Temp* = *PS* //save a copy of all positive instances
3. for i = 1:N
4. $pos = \emptyset$; //*pos*: positive subset in the mini-batch
5. if (length(*Temp*)< S_m)
6. *Temp* = random(*PS*);
7. for 1:S_m
8. randomly pick an instance t from *Temp*;
9. $pos = pos \cup t$;
10. *Temp* = *Temp* – *t;*
11. end for
12. $mb_i = x_i \cup pos$;
13. end for;
14. return *MB* to train CNNs;

In the algorithm, the parameter S (the size of the mini-batches) is dynamically defined as follows.

$$S = \lfloor m_f/M * \eta \rfloor * (1 + 1/\alpha). \quad (3)$$

Considering m_f is often small, M is set to 2 in our study. α is the positive to negative instance ratio (i.e., $\alpha = S_m / S_n$) per mini-batch. As we try to generate pseudo balanced mini-batches for each training iteration, α is set to 1 in our study. η is a compensation coefficient to accommodate a wide range of datasets with diverse data imbalance ratios. In our study, η is chosen to be 1.5 for severely imbalanced concepts and 1 otherwise. Please note that all these parameters are selected from our empirical studies. However, different values can be used according to the characteristics of the datasets.

D. Integrating CNN with low-level features

CNNs can achieve better classification performance as compared to many other classifiers, but they are computationally expensive [36], especially when applied to large multimedia datasets. To tackle this time complexity issue, we propose to extract low-level features instead of directly using raw media data such as the image RGB pixel values as input. Specifically, to work on images in our study, we carefully select five kinds of low-level features including haar [37], HOG [38], HSV [39], YCbCr [40], and CEDD [41], which are concatenated into a feature vector with 709 elements. To further reduce the computational cost, principle component analysis (PCA) is applied to reduce the size of the feature vector to 324. Because a CNN is initially designed to process an image as a 2-dimensional matrix input, here we reshape the feature vectors into 18*18 matrices. Taking into account that these generated low-level features are not always stationary, we do not apply spatial subsampling layers in our proposed framework. In order to better accommodate the reduced input feature size, the size of receptive fields (i.e., kernel size) and the number of feature maps in each convolutional layer in CNNs can be set to a relatively small number. In our proposed framework, the kernel size is set to be 3 and two convolutional layers are used with their corresponding numbers of feature maps being 6 and 9, respectively. The final output in the fully-connected layer is 2 since in our framework we only target at the binary classification problem. The small size chosen here is also helpful to reduce the computation time in the training process when compared to other work [42].

IV. EXPERIMENTAL RESULTS

To demonstrate the effectiveness of our proposed framework for multimedia data classification, it is tested on TRECVID large-size benchmark dataset with a highly imbalanced class distribution.

A. Performance Evaluation

In general, a classifier is evaluated by a confusion matrix as illustrated in Table III. The columns are the predicted class and the rows are the state of nature (actual class). In the confusion matrix, *TP* and *FP* represent the numbers of positive instances that are correctly (True Positives) or incorrectly classified (False Positives). Similarly, *TN* and *FN* indicate the numbers of negative instances being correctly (True Negatives) or incorrectly classified (False Negatives). For performance comparison, the precision and recall metrics [43] are commonly used and are derived from the confusion matrix as follows.

$$precision = \frac{TP}{TP + FP}\ ;\ recall = \frac{TP}{TP + FN}\ .$$

TABLE III. CONFUSION MATRIX

	Predicted Positive	**Predicted Negative**
State of nature Positive	True Positives (*TP*)	False Negatives (*FN*)
State of nature Negative	False Positives (*FP*)	True Negatives (*TN*)

The recall and precision goals can often be conflicting, since the increase of true positive data instances for the minority class may also increase the number of false positives, which will reduce the precision. For imbalanced data classification, the recall value is normally considered a more important criterion because it is more desirable to detect as many interesting events as possible, even at the expense of adding a reasonable number of false positives. For instance, users often want to locate all possible frauds in banking operations followed by a manual double check to root out false alarms, instead of missing true scams. In addition, *F-score*, also known as F_1 measurement or *F-value*, captures the trade-offs between precision and recall, and is considered an objective and ultimate quality metric of a classifier. It is defined as follows.

$$F - score = 2 \times \frac{precision \times recall}{precision + recall}\ . \qquad (4)$$

B. Experimental Setup

The IACC.1.B dataset is chosen from the TRECVID 2011 benchmark [44], whose semantic indexing (SIN) task aims to recognize the semantic concept contained within a video shot, which can be an essential technology for retrieval, categorization, and other video exploitations. Here, the concepts refer to high-level semantic objects such as a car, road, and tree. It has several challenges such as data imbalance, scalability, and semantic gap [45]-[46]. The IACC.I.B dataset contains approximately 8,000 Internet Archive videos (50GB, 200 hours) with creative commons licenses in MPEG-4/H.264 with the duration between 10 seconds and 3.5 minutes. Most videos have some metadata provided by the donor, e.g., title, keywords, and descriptions. These videos were collected from the Internet and were diversified in terms of the creator, content, style, production qualities, and original collection devices. The videos are segmented into a number of shots and each shot is represented by a keyframe. The shot boundaries and keyframes are also given in the dataset. The labels are provided by a collaborative annotation effort organized by NIST (National Institute of Standards and Technology). In this study, each keyframe is treated as a data instance. As discussed in [47]-[48], traditional deep learning approaches, including CNNs, often perform poorly on the TRECVID dataset due to the problem of under-fitting, huge diversity, and noisy and incomplete data annotation.

C. Experimental Results on the TRECVID dataset

The TRECVID dataset is chosen because it is largely imbalanced as some basic statistics shown in Table IV [49]. Our framework is tested on all 54 concepts that are severely imbalanced with P/N ratios between 0.0002 and 0.0005. As we discussed earlier, in imbalanced data classification, the recall metric is considered more important than precision, and the *F-score* represents the trade-off between precision and recall. Hence, as shown in Fig. 1 and Fig. 2, the *F-score* and recall values of our framework are compared with the scores from TiTech (Tokyo Institute of Technology) that achieved the best performance in TRECVID 2011 semantic indexing task [50]-[51].

TABLE IV. INFORMATION OF THE TRECVID DATASET

Data Set	**IACC.1.B**
TRECVID Year	2011
No. of Tested Concepts	54
No. of Trained Instances	262911
No. of Tested Instances	137327
P/N Ratio	[0.0002, 0.0005]

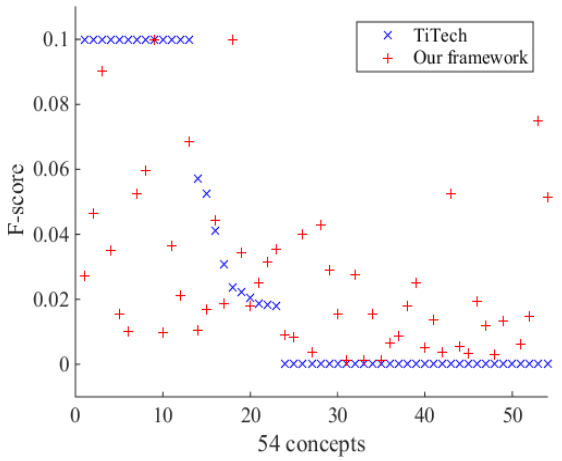

Figure 1. *F-score* comparsion for all concepts.

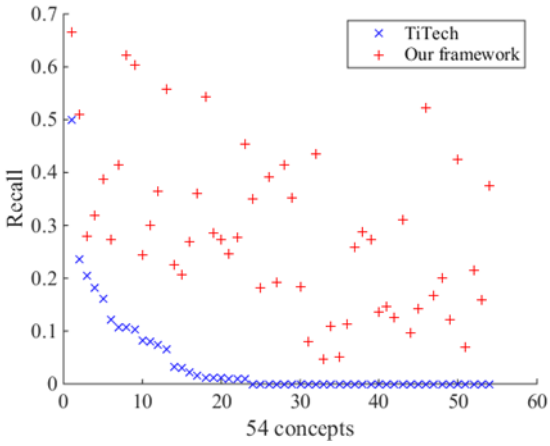

Figure 2. Recall comparsion for all concepts.

As can be seen, our F-scores are higher than those of the TiTech group (about 81.5% of the 54 concepts) and our recall values are much higher for every concept. It is also worth noting that among 54 concepts, the TiTech group can only locate zero or one true positive instance in most of the concepts; while our approach reaches about 0.29 recall value on average. This clearly demonstrates the effectiveness of our framework for imbalanced multimedia data classification.

In our second experiment, CNNs are directly applied to the TRECVID dataset without the proposed bootstrapping method. The results are shown in Fig. 3, where the x-axis indicates the number of iterations and the y-axis is the result of the prediction error rates as the convergence goes. As we can see, the error rate largely fluctuates and does not decrease during the convergence process, which means the deep learning model performs poorly on a skewed dataset. To be specific, we select three representative concepts with different levels of P/N ratios as an example (see Table V) to illustrate that CNNs are biased towards the negative instances (the majority class) and have poor classification performance on the positive instances, where all positive instances are wrongly classified as negative (i.e., with high *FN* but zero *TP*).

TABLE V. OVERFITTING FOR ORIGINAL CNN

Concepts	TP	FP	FN	TN	Total
139	0	0	73	137254	137327
43	0	0	131	137196	137327
283	0	0	27	137300	137327

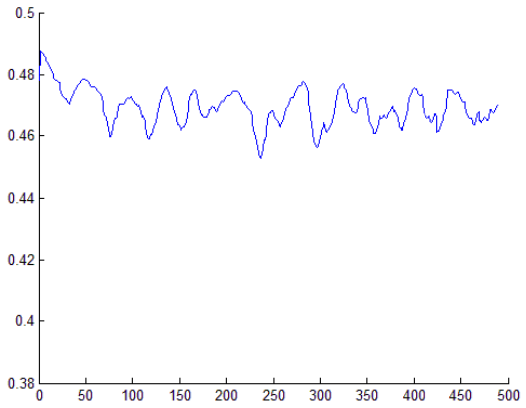

Figure 3. Error rate convergence in original CNN for an imbalanced dataset

In contrast, after applying the bootstrapping method on CNNs, the error rate can be decreased during as the convergence goes for the imbalanced dataset as shown in Fig. 4.

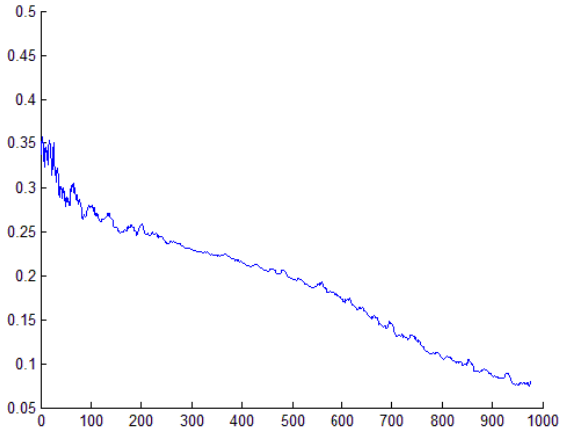

Figure 4. Error rate convergence with the bootsrapping method

In the third experiment, our framework with and without the MCA component is tested on all 54 TRECVID concepts and their average F-score values are compared in Table VI. As we can see, ADASYN+MCA that integrates ADASYN and MCA outperforms ADASYN alone for imbalanced data classification (with over 11% improvement in F-score), which shows the contributions of MCA to our framework.

TABLE VI. PERFORMANCE IMPROVEMENT AFTER MCA INTEGRATION

	ADASYN	ADASYN+MCA	Performance Improvement
F-score	0.0285	0.0317	11.2%

V. CONCLUSIONS

Deep learning has achieved great success in many research topics but little work has been done to address the issue of imbalanced data classification. It is challenging to identify as many minority instances as possible while achieving a reasonable precision performance. Most classifiers tend to be biased towards the majority class since it overshadows the minority class in the training dataset. To address this issue, in this paper, we propose an extended CNN with a bootstrapping method, where a set of pseudo balanced training mini-batches are generated and fed into CNNs for classification. To further improve the classification result, we integrate the state-of-the-art synthetic sampling method ADASYN and extend it by applying MCA to extract representative synthetic positive instances to expand the positive instance pool. This process changes the data distribution to make the training dataset less imbalanced and supplies the CNNs with more minority class samples. By using the TREVID dataset as the testbed, the experimental results demonstrate the effectiveness of our framework in classifying multimedia data with a highly skewed data distribution.

ACKNOWLEDGMENT

For Shu-Ching Chen, this research is partially supported by NSF HRD-0833093, CNS-1126619, and NSF CNS-1461926.

REFERENCES

[1] X. Huang, S.-C. Chen, M.-L. Shyu, and C. Zhang, "User Concept Pattern Discovery Using Relevance Feedback and Multiple Instance Learning for Content-Based Image Retrieval," Proceedings of the Third International Workshop on Multimedia Data Mining (MDM/KDD'2002), in conjunction with the 8th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, pp. 100-108, July 23, 2002, Edmonton, Alberta, Canada.

[2] S.-C. Chen, S. Sista, M.-L. Shyu, and R. L. Kashyap, "Augmented Transition Networks as Video Browsing Models for Multimedia Databases and Multimedia Information Systems," Proceedings of the 11th IEEE International Conference on Tools with Artificial Intelligence (ICTAI'99), pp. 175-182, November 9-11, 1999, Chicago, IL, USA.

[3] S.-C. Chen, S. H. Rubin, M.-L. Shyu, and C. Zhang, "A Dynamic User Concept Pattern Learning Framework for Content-Based Image Retrieval," IEEE Transactions on Systems, Man, and Cybernetics: Part C, Vol. 36, Issue 6, pp. 772-783, November 2006.

[4] S.-C. Chen, S. H. Rubin, M.-L. Shyu, and C. Zhang, "A Dynamic User Concept Pattern Learning Framework for Content-Based Image Retrieval," IEEE Transactions on Systems, Man, and Cybernetics: Part C, Vol. 36, Issue 6, pp. 772-783, November 2006.

[5] X. Li, S.-C. Chen, M.-L. Shyu, and B. Furht, "Image Retrieval by Color, Texture, and Spatial Information," Proceedings of the 8th International Conference on Distributed Multimedia Systems (DMS'2002), pp. 152-159, September 26-28, 2002, San Francisco Bay, California, USA.

[6] M.-L. Shyu, C. Haruechaiyasak, and S.-C. Chen, "Category Cluster Discovery from Distributed WWW Directories," Journal of Information Sciences, special issue on Knowledge Discovery from Distributed Information Sources, Vol. 155, Issues 3-4, pp. 181-197, October 2003.

[7] X. Li, S.-C. Chen, M.-L. Shyu, and B. Furht, "An Effective Content-Based Visual Image Retrieval System," Proceedings of the 26th IEEE Computer Society International Computer Software and Applications Conference (COMPSAC), pp. 914-919, August 26-29, 2002, Oxford, England.

[8] M.-L. Shyu, S.-C. Chen, M. Chen, C. Zhang, and K. Sarinnapakorn, "Image Database Retrieval Utilizing Affinity Relationships," Proceedings of the First ACM International Workshop on Multimedia Databases (ACM MMDB'03), pp. 78-85, November 7, 2003, New Orleans, Louisiana, USA.

[9] X. Chen, C. Zhang, S.-C. Chen, and S. H. Rubin, "A Human-Centered Multiple Instance Learning Framework for Semantic Video Retrieval," IEEE Transactions on Systems, Man, and Cybernetics: Part C, Vol. 39, Issue 2, pp. 228-233, March 2009.

[10] M.-L. Shyu, S.-C. Chen, and R. L. Kashyap, "Generalized Affinity-Based Association Rule Mining for Multimedia Database Queries," Knowledge and Information Systems (KAIS): An International Journal, vol. 3, no. 3, pp. 319-337, August 2001.

[11] Q. Zhu, L. Lin, M.-L. Shyu, and S.-C. Chen, "Effective Supervised Discretization for Classification based on Correlation Maximization," Proceedings of the 12th IEEE International Conference on Information Reuse and Integration (IRI 2011), Las Vegas, Nevada, USA, pp. 390-395, August 3-5, 2011.

[12] H.-B. He and E.A.Garcia, "Learning from Imbalanced Data," IEEE Transactions on Knowledge and Data Engineering, Vol. 21, No. 9, pp. 1263-1284. September 2009.

[13] S.-C. Chen, M.-L. Shyu, and C. Zhang, "An Intelligent Framework for Spatio-Temporal Vehicle Tracking," Proceedings of the 4th International IEEE Conference on Intelligent Transportation Systems, pp. 213-218, August 25-29, 2001, Oakland, California, USA.

[14] L. Lin and M.-L. Shyu, "Weighted Association Rule Mining for Video Semantic Detection," International Journal of Multimedia Data Engineering and Management (IJMDEM), Vol. 1, No. 1, pp. 37–54, 2010.

[15] S.-C. Chen, M.-L. Shyu, C. Zhang, and R. L. Kashyap, "Identifying Overlapped Objects for Video Indexing and Modeling in Multimedia Database Systems," International Journal on Artificial Intelligence Tools, Vol. 10, No. 4, pp. 715-734, December 2001

[16] X. Chen, C. Zhang, S.-C. Chen, and M. Chen, "A Latent Semantic Indexing Based Method for Solving Multiple Instance Learning Problem in Region-based Image Retrieval," Proceedings of the IEEE International Symposium on Multimedia (ISM 2005), pp. 37-44, December 12-14, 2005, Irvine, California, USA.

[17] S.-C. Chen, S. Sista, M.-L. Shyu, and R. L. Kashyap, "Indexing and Searching Structure for Multimedia Database Systems," Proceedings of the IS&T/SPIE conference on Storage and Retrieval for Media Databases 2000, pp. 262-270, January 23-28, 2000, San Jose, CA, USA.

[18] S.-C. Chen, "Multimedia Databases and Data Management: A Survey," International Journal of Multimedia Data Engineering and Management, Vol. 1, No. 1, pp. 1-11, January-March 2010.

[19] S.-C. Chen, S. Sista, M.-L. Shyu, and R. L. Kashyap, "Indexing and Searching Structure for Multimedia Database Systems," Proceedings of the IS&T/SPIE conference on Storage and Retrieval for Media Databases 2000, pp. 262-270, January 23-28, 2000, San Jose, CA, USA.

[20] C. Chen and M.-L. Shyu, "Integration of Semantics Information and Clustering in Binary-class Classification for Handling Imbalanced Multimedia Data," Edited by Tansel Ozyer, Keivan Kianmehr, Mehmet Tan, and Jia Zeng, Information Reuse and Integration in Academia and Industry, Chapter 14, Springer Verlag, 2013.

[21] C. Chen and M.-L. Shyu, "Clustering-based Binary-class Classification for Imbalanced Data Sets," Proceedings of the 12th IEEE International Conference on Information Reuse and Integration, pp. 384-389, Las Vegas, Nevada, USA, August 2011.

[22] L. Lin, G. Ravitz, M.-L. Shyu, and S.-C. Chen, "Video Semantic Concept Discovery using Multimodal-based Association Classification," Proceedings of the IEEE International Conference on Multimedia & Expo, pp. 859-862, Beijing, China, July 2-5, 2007.

[23] M.-L. Shyu, S.-C. Chen, Q. Sun, and H. Yu, "Overview and Future Trends of Multimedia Research for Content Access and Distribution," International Journal of Semantic Computing (IJSC), Vol. 1, No. 1, pp. 29-66, March 2007.

[24] L.-S. Chen, C.-C. Hsu, and Y.-S. Chang, "MDS: A Novel Method for Class Imbalance Learning," Proceedings of the International Conference on Ubiquitous Information Management and Communication, pp. 544-549, Janurary 2009.

[25] J. Wan, D. Wang, S. C. H. Hoi, P. Wu, J. Zhu, Y. Zhang, and J. Li, "Deep Learning for Content-Based Image Retrieval: A Comprehensive Study," Proceedings of the ACM International Conference on Multimedia, pp. 157-166, Novemember 2014.

[26] H.-B. He, Y. Bai, E.A. Garcia and S.-T. Li, "ADASYN: Adaptive Synthetic Sampling Approach for Imbalanced Learning," Proceedings of the International Joint Conference on Neural Networks, pp.1322-1328, June 2008

[27] L. Zhang and W. Wang, "A Re-sampling Method for Class Imbalance Learning with Credit Data," Proceedings of the 2011 International Conference on Information Technology, Computer Engineering and Management Sciences, pp. 393-397, September 2011.

[28] N. V. Chawla and K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "SMOTE: Synthetic Minority Overbootstrapping Technique," Journal of Artijcial Intelligence Research, 16, pp. 321-357, 2002.

[29] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner. "Gradient-based Learning Applied to Document Recognition," Proceedings of the IEEE, Vol. 86, No. 11, pp.2278-2324, November 1998.

[30] E. R. Kandel, "An Introduction to the Work of David Hubel and Torsten Wiesel," The Journal of Physiology 587 (Pt 12), pp. 2733–2741, April 2009.

[31] J. Bouvrie, "Notes on Convolutional Neural Networks.," Technical Report, 2006.

[32] P. Swietojanski, A. Ghoshal, and S. Renals, "Convolutional Neural Networks for Distant Speech Recognition," IEEE Signal Processing Letters, Vol. 21, No. 9, pp. 1120-1124, September 2014.

[33] X. Chen, S. Xiang, C.-L. Liu, and C.-H. Pan, "Vehicle Detection in Satellite Images by Hybrid Deep Convolutional Neural Networks," IEEE Geoscience and Remote Sensing Letters, Vol. 11, No. 10, pp. 1797-1801, Octorber 2014.

[34] Q. Mao, M. Dong, Z. Huang, and Y. Zhan, "Learning Salient Features for Speech Emotion Recognition Using Convolutional Neural Networks," IEEE Transactions on Multimedia, Vol. 16, No. 8, pp. 2203-2213, December 2014.

[35] Y. Yan, Y. Liu, M.-L. Shyu, and M. Chen, "Utilizing Concept Correlations for Effective Imbalanced Data Classification," Proceedings of the 2014 IEEE 15th International Conference on Information Reuse and Integration, pp. 561-568, August 13-15, 2014.

[36] A. Karpathy, G. Toderici, S. Shetty, T. Leung, R. Sukthankar, and L. Fei-Fei, "Large-Scale Video Classification with Convolutional Neural Networks," Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 1725-1732, June 2014.

[37] D. Verma and V. Maru. "An Efficient Approach for Color Image Retrieval using Haar Wavelet," Proceedings of the IEEE International Conference on In Methods and Models in Computer Science, pp. 1–5, 2009.

[38] N. Dalal and B. Triggs, "Histograms of Oriented Gradients for Human Detection," Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 1, pp. 886-893, June 2005.

[39] J.C. Femiani and A.Razdan, "Interval HSV: Extracting ink annotations" Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 2520-2527, June 2009

[40] S. Sural, G. Qian, and S. Pramanik, "Segmentation and Histogram Generation using the HSV Color Space for Image Retrieval," Proceedings of the International Conference on Image Processing (ICIP), pp. 589-592, 2002.

[41] S. A. Chatzichristofis and Y. S. Boutalis, "CEDD: Color and Edge Directivity Descriptor: A Compact Descriptor for Image Indexing and Retrieval," Proceedings of the 6th International Conference on Computer Vision Systems, pp. 312–322, Berlin, Heidelberg, 2008.

[42] K. He, X. Zhang, S. Ren, and J. Sun, "Spatial Pyramid Pooling in Deep Convolutional Networks for Visual Recognition," Proceedings of the European Conference on Computer Vision, pp. 346-361, Zurich, Switzerland, September 6-12, 2014.

[43] M. Buckland and F. Gey, "The Relationship Between Recall and Precision," Journal of the American Society for Information Science, Vol. 45, No. 1, pp. 12-19, January 1999.

[44] A. F. Smeaton, P. Over, and W. Kraaij, "Evaluation Campaigns and TRECVid," Proceedings of the 8th ACM International Workshop on Multimedia Information Retrieval, pp. 321-330, 2006.

[45] T. Meng and M.-L. Shyu, "Leveraging Concept Association Network for Multimedia Rare Concept Mining and Retrieval," Proceedings of the 2012 IEEE International Conference on Multimedia and Expo, pp. 860-865, Washington DC, USA, July 2012.

[46] T. Meng and M.-L. Shyu, "Automatic Annotation of Drosophila Developmental Stages using Association Classification and Information Integration," Proceedings of the 2011 IEEE International Conference on Information Resue and Integration, pp. 142–147, Las Vegas, Nevada, August 2011.

[47] Y. Sun, T. Osawa, K. Sudo, Y. Taniguchi, H. Li, Y. Guan, and L. Liu, "TRECVid 2013 Semantic Video Concept Detection by NTT-MD-DUT," TRECVID 2013, November 26 – 28, 2013.

[48] C.G.M. Snoekyz, K.E.A. van de Sandeyz, D. Fontijnez, A. Habibiany, M. Jain, S. Kordumovay, Z. Liy, M. Mazloomy, S.L. Pinteay, R. Taoy, D.C. Koelmayz, and A.W.M. Smeulders, "MediaMill at TRECVID 2013: Searching Concepts, Objects, Instances and Events in Video," TRECVID 2013, November 26 – 28, 2013.

[49] A. F. Smeaton, P. Over, and W. Kraaij, "Evaluation Campaigns and TRECVid," Proceedings of the 8th ACM International Workshop on Multimedia Information Retrieval, pp. 321-330, 2006.

[50] N. Inoue, T. Wada, Y. Kamishima, K. Shinoda, and S. Sato, "TokyoTech+Canon at TRECVID 2011," Proceedings of the TRECVID Workshop 2011, December 5, 2011.

[51] N. Inoue and K. Shinoda, "A Fast and Accurate Video Semantic-Indexing System Using Fast MAP Adaptation and GMM Supervectors," IEEE Transactions on Multimedia, Vol. 14, No. 4-2, pp. 1196-1205, 2012.