International Journal of Semantic Computing
© World Scientific Publishing Company

Correlation-Assisted Imbalance Multimedia Concept Mining and Retrieval

Yilin Yan* and Mei-Ling Shyu†

Department of Electrical and Computer Engineering
University of Miami
Coral Gables, Florida 33124, USA
**y.yan4@umiami.edu*
†*shyu@miami.edu*

In the past decades, we have witnessed an explosion of multimedia data, especially with the development of social media websites and blooming popularity of smart devices. As a result, multimedia semantic concept mining and retrieval whose objective is to mine useful information from the large amount of multimedia data including texts, images, and videos has become more and more important. The huge amount of multimedia data and the semantic gap between low-level features and high-level semantic concepts have made it even more challenging. To address these challenges, the correlations among the classes can provide important context cues to help bridge the semantic gap. Meanwhile, many real-world datasets do not have uniform class distributions while the minority instances actually represent the concept of interests, like frauds in transactions, intrusions in network security, and unusual events in surveillance. Despite extensive research efforts, imbalanced concept retrieval remains one of the most challenging research problems in multimedia data mining. Different from existing frameworks regarding concept correlations among labels, this paper presents a novel concept correlation analysis model using the correlation between the retrieval scores and labels. Experimental results on the TRECVID benchmark datasets demonstrate that the proposed framework can enhance imbalanced concept mining and retrieval even with trivial scores from the minority class.

Keywords: Imbalanced data; Multimedia big data; Multimedia semantic retrieval; Rare class mining; Concept correlation; Information integration

1. Introduction

The class imbalance problem has attracted significant research efforts in data mining and information retrieval [1][2][3][4][5][6]. Large amounts of data have skewed class distributions since the events of interests occur infrequently in many real-world big datasets. In such datasets, classes with fewer data instances are called minority classes; while those have more data instances are defined as majority classes. Most classifiers often bias towards the majority classes since they are usually modeled by exploring data statistics and thus can hardly retrieve correct results from the minority classes. However, the minority instances are usually the most important ones in the fields of risk management, rare disease in diagnosis, metal fatigue detection, as well as multimedia concept retrieval which is one of the centric research tasks in

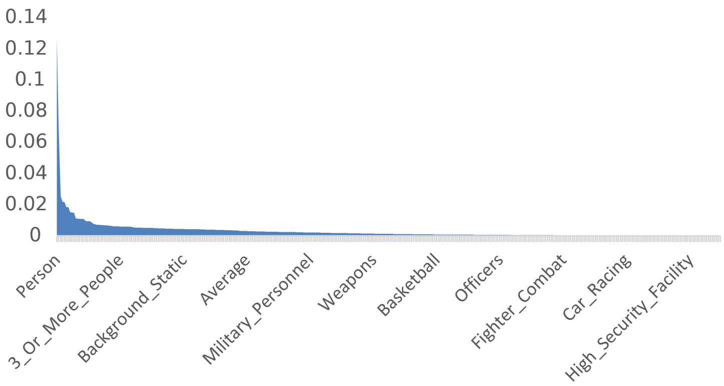

Fig. 1. Positive to Negative (P/N) Ratios for some rare concepts

content-based information retrieval [7][8][9][10][11][12][13].

Although researches have paid extensive efforts on the class imbalance problem, rare concept retrieval remains one of the most challenging problems in multimedia data [14][15][16][17][18][19][20]. Many concepts are often correlated, either positively or negatively. Some concepts co-occur rarely like cow and sea; while others co-occur more frequently such as bird and sky. Such correlations can provide important context cues to help detect the concepts [21][22][23][24][25][26]. While inter-concept correlations have been recently used to tackle the issue, the very small number of training instances in the minority class makes the task of correlation detection hard and often lead to unsatisfied concept retrieval results. Different from those enhancement models that only consider the correlations among concepts, we present a very different correlation analysis strategy considering the correlation between concept labels and retrieval scores. Even with trivial scores from minority classes, the proposed framework can enhance rare concept retrieval.

The main contribution of this paper is the design of an efficient multimedia rare concept retrieval model. This paper focuses on efficient imbalance concept mining in a large-scale dataset, namely the TREC Video Retrieval (TRECVID) dataset [27] which includes a lot of videos collected from the Internet and other sources by National Institute of Standards and Technology (NIST). Many concepts are considered imbalanced as shown in Figure 1). As mentioned previously, some concepts are extremely rare like “Cigar Boats” which contains only four keyframes (video shots), making semantic information retrieval a big problem. The average P/N (positive to negative) ratio of the TRECVID data is only 0.003. By constructing a semantic concept hierarchy and using concept correlations, a novel imbalanced concept retrieval model is proposed in this paper. Experimental results on TRECVID 2015 semantic indexing (SIN) data set demonstrate that the proposed framework gives promising performance, comparing to several state-of-the-art approaches.

The rest of this paper is organized as follows. In the next section, related work on rare class mining is introduced and various types of correlations are discussed as well. In Section 3, a hierarchy is built using the inter-concept correlations. Section 4 describes a novel idea of enhancing imbalanced concept detection using the correlation between the retrieval scores and labels. Section 5 shows how to setup the framework and compares the results of the proposed system on the TRECVID dataset. Finally, Section 6 draws the conclusion and identifies future research directions.

2. Related work

In this section, we first introduce some kinds of correlation coefficients for different types of data. Next, some recent approaches on imbalance data classification are discussed and divided into two directions. We also include how to build the hierarchies for classes as part of the related work.

2.1. *Correlation coefficients*

In general, the correlation coefficient, known as the cross-correlation coefficient, is a quantity that gives the statistical relationships between two or more random variables or observed data values. Based on the nature of the input data, inter-concept correlations can be divided into four different kinds including nominal, ordinal, interval, and ratio.

When measuring using a nominal scale, one simply names or categorizes the responses. The input data for the nominal scale are put into classes without any structure or order. Considering a dataset of hair colors including labels brown, black, gray, red, etc. There is no distance between brown and black, nor a distance between gray and red. A sub-type of nominal scales with only two categories is called “dichotomous” (like male and female). All categories in nominal scales have no overlap and none of them has any numerical significance. Therefore, they can be also seen as kinds of names or yes/no labels.

Similar with nominal scales, ordinal scales are typically measures of non-numeric concepts like satisfaction, happiness, discomfort, etc. However, the differences between them is not really known. For instance, it is typically unknown whether the difference between “Very Happy” and “Happy” is the same as the difference between “So-so” and “Unhappy”. Therefore, ordinal scales can only interpret a gross order but not the relative positional distances. Besides, the best way to determine the central tendency on a set of ordinal data is to use the mode or median rather than the mean from an ordinal set since the “mean” value is always undetermined.

The third, interval scales, which we are most familiar with, are numeric scales in which we know not only the order, but also the exact differences between the values and thus the realm of statistical analysis on such data opens up, which is to say the intervals having the same interpretation throughout. The Fahrenheit temperature is considered the classic example data in this category. While 10 degrees plus 10 degrees is 20 degrees, 20 degrees is not twice as hot as 10 degrees since Fahrenheit temperature do not have a true zero point, even if one of the scaled values happens to carry a zero value (0 degree, not absolute zero). Therefore, with interval data, we can add and minus, but cannot multiply or divide.

Lastly, ratio scales are kind of ultimate nirvana because they tell us about the order, the exact value between units, as well as an absolute zero which allows for a wide range of both descriptive and inferential statistics to be applied. Therefore, different from interval scales, the multiply and divide operators can also be applied. One common example of ratio scales is the height, since we can always say one object is twice as tall as another. The summary of the operations and scale measures is shown in Table 1.

Table 1. Summary of Data Types and Scale Measures

Operations	Nominal scales	Ordinal scales	Interval scales	Ratio scales
Frequent distribution	Yes	Yes	Yes	Yes
Mode	Yes	Yes	Yes	Yes
Median	Yes	Yes	Yes	Yes
Mean	No	Yes	Yes	Yes
Plus	No	No	Yes	Yes
Minus	No	No	Yes	Yes
Multiple	No	No	No	Yes
Divide	No	No	No	Yes

2.2. *Imbalanced data classification*

Recently, imbalanced data classification techniques fall into two categories including sampling-based and algorithm-based. Sampling-based approaches are the most popular classification algorithms for imbalanced data sets. Among them, downsampling (undersampling) and oversampling methodologies have received significant research efforts to counter the classification of imbalanced datasets and presented the viewpoints on the usefulness of undersampling versus oversampling [28], though sometimes they are conflicting. While the sampling ideas are straightforward, they have proven good performance on imbalanced data classification.

Downsampling is to select a part of the majority data instances to build a model with a similar number of positive samples. The main advantage of undersampling is its efficiency as it uses only a subset of the data instances in the majority class; while many data instances in the majority class are ignored, but this may lead to information lost. To overcome the disadvantage of downsampling, Liu et al. proposed two algorithms to overcome this deficiency [29]. The first one is "Easy Ensemble" which trains classification models using several sample subsets from the majority class, and then integrates the outputs of those models to produce the final predication results. The second promoted one is "Balance Cascade" which trains the classification models sequentially. In each step, the majority class data instances that are correctly classified by the current trained models are removed from the next round. Although downsampling is somewhat efficient as it uses only a subset of the majority class, many data instances in the majority class are ignored and may result in the loss of information.

Comparatively, the idea of oversampling is to somehow generate more positive data instances to make the data set balanced. One easy way is to simply copy data instances in the minority class, which may lead to overfitting. Zhang et al. [30] presented an improved oversampling approach based on the synthetic minority oversampling technique (SMOTE). First, the data distribution of the minority class is used to estimate whether different types of data instances are overlapped. In the next step, synthetic data instances are generated in different classes when the classes overlap significantly with each other. In addition, the weights are increased for those positive samples that are far from the borderline. Though SMOTE is an enhanced oversampling approach which could generate data instances not existing in the original minority class, overfitting remains as a potential problem in oversampling.

Another direction of solving imbalanced data classification is to use the algorithm-based approaches. Researchers use them to optimize the performance of learning algorithms on unseen data to address the class imbalance problem. One algorithm-based approach is cost-sensitive learning which tries to maximize the loss functions associated with a dataset to improve the classification performance. It is motivated by the observation that most real-world applications do not have uniform costs for misclassifications. The actual costs associated with each kind of errors are typically unknown, so these methods need to determine the cost matrix based on the data and apply it to the learning stage. Shifting the bias of a machine to favor the minority class is a similar idea with cost-sensitive learning [31].

The algorithm of GASEN (Genetic Algorithm based Selective Ensemble Network) has been proven to be a very effective way to select a subset of neural networks to form an ensemble classifier or a regressor of an enhanced generation ability. Che et al. provided an improved solution of GASEN to handle the class-imbalance problem and tested GASEN on dozens of datasets to find that there is some potential for improving GASENs performance on class-imbalance learning [32]. Machine learning algorithms, such as genetic programming (GP), can also generate biased classifiers when the data sets are imbalanced. Bhowan et al. used new fitness functions in the GP learning process and empirically showed a better performance by the evolved classifiers on both minority and majority classes [33]. Though these studies have shown their potentials in improving the classification performance on imbalanced data, they are far from extensive or systematic.

2.3. *Hierarchical models*

Many research efforts have been paid on organizing the hierarchies for semantic concept retrieval and event detection. Most of them use inter-concept correlations to build the hierarchies. Wang et al. [34] proposed a hierarchical context model to systematically integrate feature level context, semantic level context, and prior level context for accurate and robust event recognition in surveillance videos. A comprehensive model that can integrate contexts from all three levels simultaneously was built. In [35], the authors presented a large-scale video event classification system with a large number of event categories mined automatically from YouTube video titles and descriptions using Part-of-Speech parsing tools, with constraints derived from WordNet hierarchies. To solve the problem of multi-class object detection, the authors proposed a boosted multi-class object cascade that only splits one class object from the upper-level cascade when building the sub-cascades [36], which reduces the number of classifiers in each stage. Vreeswijk et al. [37] analyzed the differences between the images labeled at varying levels of abstraction and the union of their constituting leaf nodes.

Recently, some researchers find that inter-concept correlations can help re-rank the concept detection scores on event detection. The selection of event-specific concepts based on the similarity to a textual event description had shown to yield effective event detection results without positive examples [38]. Tao et al. [39] showed that inter-concept associations including both positive and negative correlations can be used to bridge the semantic gap and enhance the performance of semantic concept detection in multimedia data [40][41][42][43]. The concept-concept association information integration and multi-model collaboration framework were proposed to enhance high-level information retrieval from multimedia big data.

3. Building hierarchies for datasets

3.1. *Conditional probability calculation*

Although the inter-concept connection information has been proposed to enhance semantic concept retrieval results, most of them utilize the hierarchical relationship from the data provider [44] for combining the classes to generate reorganized hierarchies. For instance, if a dataset contains labels “apple”, “banana”, and “fruit”, the data provider may give a note “apple imply fruit”. In such case, the data instances with the label “apple” would be automatically added to the label “fruit”. However, many other kinds of relationships are not that straightforward and relationships generated manually may lead to biases and not suitable for big datasets.

In this paper, we first build a hierarchical model for all concepts based on conditional probabilities generated from the training set. Here we define C_{parent} as a parent concept and C_{child} as a child concept. Let $P(.)$ be the probability, then C^+_{parent} denotes the positive collection of C_{parent}; whereas C^+_{child} represents the positive collection of C_{child}. If C_{parent} is the parent of C_{child}, the appearance of C_{child} should imply the appearance of C_{parent}. As an example, if a video shot contains the concept “car”, it definitely includes the concept “vehicle” as well, unless the ground truth is incorrect. In this example, “car” is a child concept while “vehicle” is a parent concept. The probability of C_{parent} appearing increases if C_{child} appears, and the probability of C_{parent} appearing decreases if C_{child} does not appear. This

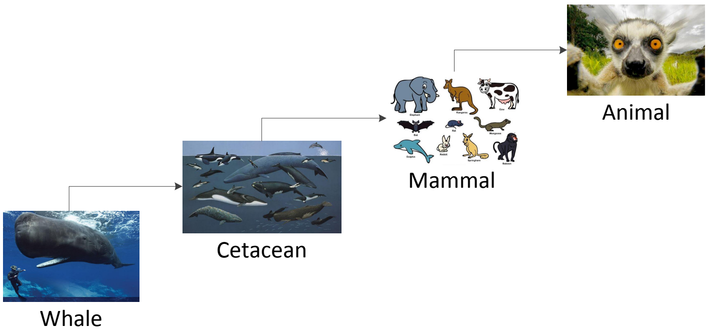

Fig. 2. Parent/Child relationship examples 1.

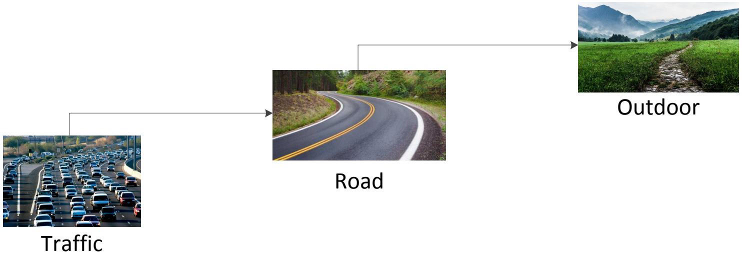

Fig. 3. Parent/Child relationship examples 2.

conditional probability can be computed by Equation (1).

$$P(C^+_{parent}|C^+_{child}) = \frac{P(C^+_{parent} \text{ and } C^+_{child})}{P(C^+_{child})}. \tag{1}$$

3.2. *Bottom-up organization*

In real-world, some concept pairs have the parent-child relationship (like “sky” and “sun”). This kind of inter-concept relationships should also be considered. In addition, since the concept labels in multimedia datasets are usually manually decided by the volunteers or by some automatic labeling techniques, the ground truth is not always correct. Therefore, a threshold of 0.9 is set to determine whether two concepts have the parent/child relationship, which is represented as $P(C^+_{parent}|C^+_{child})$ > 0.9.

Next, the hierarchy model of all concepts is built from the leaf nodes (in a bottom-up manner) using all the parent-child concept pairs generated and filtered. If a concept has no child but at least one parent, it is considered as a leaf node and is added to the initial model. The following step shows the example of including the “direct” parent nodes for the “whale” leaf node. A whale is a cetacean, a mam-

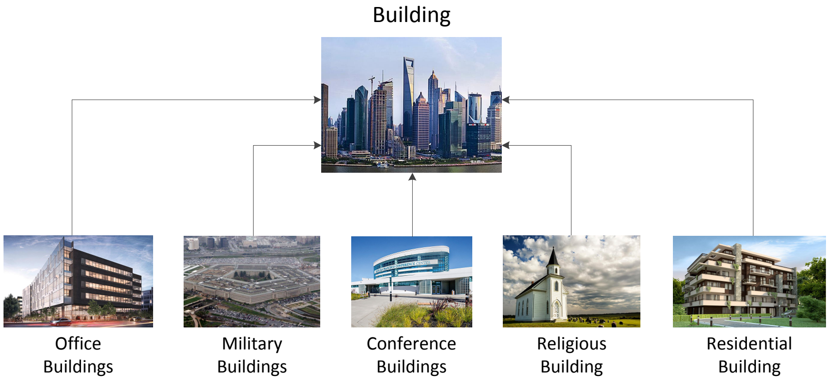

Fig. 4. Siblings relationship examples 1.

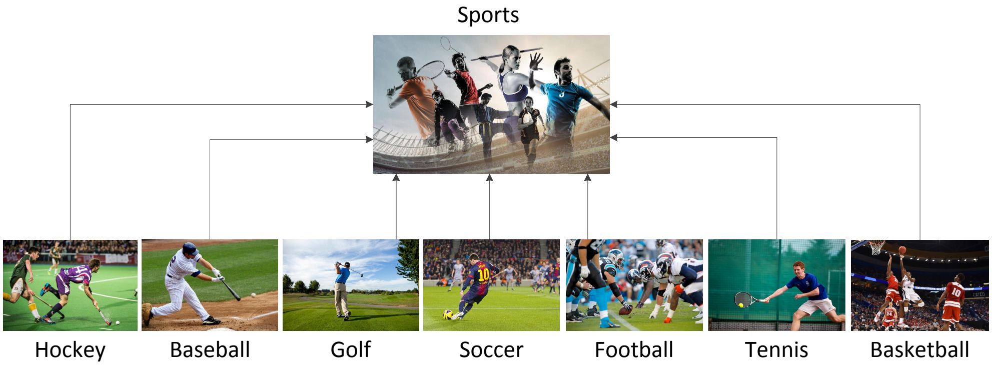

Fig. 5. Siblings relationship examples 2.

mal, and an animal as well. With the fact that the appearance of a whale implies the appearance of a cetacean and "cetacean" implies "animal", these two concept pairs also have the parent-child relationship. Thus, "whale" is first included as a child node and then followed by "cetacean", "mammal" , and "animal". If a parent concept has no parent like "animal" in this case, it will be finally considered as a root (head) node. These operations are shown in Figure 2 and Figure 3.

After finding out all the qualified parent-child concept pairs, we can combine the branches into a tree and thus find the siblings of the child concepts as given in Figure 4 and Figure 5. Different tree structures would be generated from different datasets, even from different subsets of a dataset. In the aforementioned example, if the concept "mammal" is removed, "animal" could be the direct parent of "cetacean" in the updated hierarchy. In general, the more concepts included, the more complete the model would be. Though the hierarchy model can never be perfect, it is suitable for the particular dataset on which it based.

4. Prediction score enhancement for rare concept retrieval

4.1. *Score-based correlation generation*

As mentioned in related work, most previous research including the aforementioned conditional probability approaches calculates the inter-concept correlations and builds the hierarchical structures using the label information in the training data, i.e., the appearance or non-appearance of the concepts. One main problem of using such information to leverage the retrieval scores is the correlation coefficients among rare concepts, and correlation coefficients between imbalanced concepts and balanced concepts are usually weak. Suppose that we calculate the correlation between a common concept with 10,000 instances and a rare concept with 10 instances, the correlation coefficient will be small and even one wrong label for the rare concept in the training set will lead to a big mistake in inter-concept correlation calculation and thus cause wrong results.

Another issue is that high correlations between concepts do not necessarily lead to the high correlation between the concepts and the detection (prediction) scores, especially for rare concepts since the quality of scores from imbalanced concepts is often worse than those from the balanced concepts. This is caused by the nature of the original dataset (with a skewed distribution) and directly using rare concepts' correlation information for score integration may even downgrade the original results. For instance, the concept "hurricane" should have a positive correlation with the concept "disaster". However, with the bad prediction scores, the concept "hurricane" does not really help the retrieval of the concept "disaster" in the imbalanced dataset.

There are only 6 out of the total of 137,272 video shots that include the concept "cow" in the TRECVID dataset. This raises the third issue: the detection scores of rare concepts themselves can be relatively imprecise. Most of the classifiers cannot get acceptable prediction scores for these rare concepts albeit with such a big training data set. To solve these three issues, we propose a model to integrate the prediction scores of the rare concepts using the Pearson correlation coefficients from both the label and score information for score enhancement in this paper.

The Pearson product-moment correlation coefficient [45], denoted by ρ (or r), measures the strength of a linear association between two variables X and Y, and is widely used as a measure of the degree of the linear dependence between X and Y. It attempts to draw a line of best fit through the data of two variables, and ρ (or r) indicates how far away all these data points are to this line of best fit. The ρ (or r) values are between +1 and -1 (inclusive), where +1 is a total positive correlation, 0 is no correlation, and -1 is a total negative correlation. Let $cov(X, Y)$ and E be the covariance and expectation of X and Y, σ_X and σ_Y be the standard deviations of X and Y, and μ_X and μ_Y be the mean values of X and Y. For a population, we

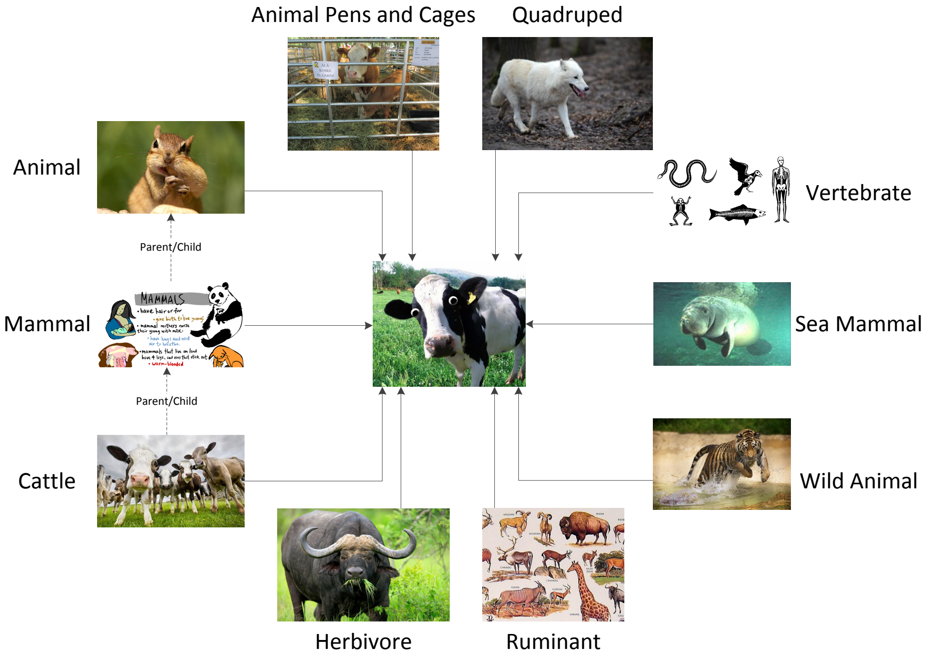

Fig. 6. Top ten related concepts that support the rare concept “cow”.

have the following:

$$\rho_{X,Y} = \frac{cov(X,Y)}{\sigma_X \sigma_Y} = \frac{E[(X-\mu_X)(Y-\mu_Y)]}{\sigma_X \sigma_Y};\ \text{where} \tag{2}$$

$$cov(X,Y) = E\{[x-E(x)][(y-E(y)]\} = E(xy) - E(x)E(y);$$

$$\sigma_X = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(x_i-\mu_X)^2};$$

$$\sigma_Y = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(y_i-\mu_Y)^2};$$

$$\mu_X = \frac{1}{N}\sum_{i=1}^{N}x_i;$$

$$\mu_Y = \frac{1}{N}\sum_{i=1}^{N}y_i.$$

Here we define C_T as the label information of an imbalanced (target) concept, and let S_R be the prediction score of a support (related) concept which can be either a balanced concept or imbalanced one. Take “cow” as a target concept. In order to enhance the prediction score of the rare concept “cow”, all $\rho_{(C_T,S_R)}$ are calculated and ranked. In the equation, T is the concept "cow" and $R = 1, 2, \cdots, N$ and N is the number of concepts. The top ten related concepts are shown in Figure 6, which means the prediction scores of these concepts are helpful to enhance the prediction score of the concept "cow".

As shown in Figure 6, the top ten related concepts are "Herbivore", "Ruminant", "Mammal", "Quadruped", "Wild Animal", "Vertebrate", "Animal", "Animal Pens And Cages", "Sea Mammal", and "Cattle", respectively. Clearly, most of them are reasonable at the first glance, expect "Sea Mammal". Nevertheless, the shapes of some sea mammals are similar to those of the cows. Especially, one common kind of sea mammal, manatee, is also known as "sea cow". This highlights another advantage of the proposed framework, which can find the potentially related concepts. Figure 6 also implies that the prediction score of the concept "cow" itself is imprecise and thus will not be integrated for the enhancement.

4.2. *Negative-related concepts*

After ranking top ten related concepts by their correlation values and building a hierarchy as shown in Figure 6, it can be further expanded using the hierarchical models built in Section 3. For a target concept C_T, if a related concept C_R is connected to it, its parent will also be added to the model. In this example, since "Quadruped" is connected to "cow", "Animal" would be included as well. However, since "Animal" is already included based on the ranked scores, we don't need to add it again as shown in Figure 6. In this paper, C_T is a rare concept. Afterward, the scores of the top ten related concepts are used to train an integration model using a discriminant analysis classifier.

As discussed earlier, some concepts such as "sky" and "shark" rarely co-occur, which can also provide important context cues to help detect the concepts. Take the aforementioned example, the top ten concepts that have a negative relationship with "cow" are "Hospital", "Bomber Bombing", "Fear", "Factory", "Sports Car", "Disgust", "Handshaking", "Airplane Landing", "Black Frame", and "Network Logo". Here, these 10 concepts are not simply irrelevant to "cow" as they look like, but they also have relatively strong negative relationships. That is, if "Hospital" appears in a testing frame, "cow" is very unlikely to appear in the same frame. Therefore, the opposite numbers of scores from those negative-related concepts can be integrated in the enhancement framework.

4.3. *Score integration*

To train the integration model, 5 different kinds of popular algorithms are used, including Support Vector Machine (SVM) [46][47], Naive Bayes (NB) [48], Random Forest (RF) [49][50], Logistic Regression (LR) [4][51][52], and Discriminant Analysis Classifier (DAC) [51][52][4].

A Support Vector Machine (SVM) is a discriminative classifier formally defined by a separating hyperplane to classify the dataset so that the geometric margin is

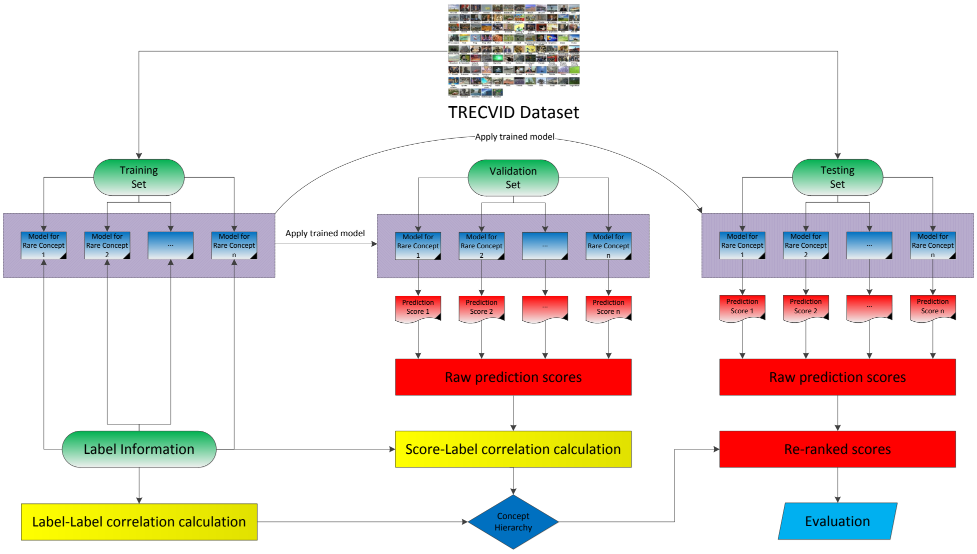

Fig. 7. The proposed framework.

maximized. A “Naive Bayes” (NB) is a classification technique based on the Bayes’ Theorem with an assumption of independence among predictors. A Random Forest (RF) is a meta estimator that fits a number of decision tree classifiers on various sub-samples of the dataset and uses averaging to improve the predictive accuracy and control over-fitting. Logistic Regression (LR) is another predictive analysis and a kind of generalized linear model. It can be used to conduct when the dependent variable is binary just like in our case. Instead of just predicting binary-valued labels in linear regression, logistic regression uses a different hypothesis class to predict the probability that a given example belongs to the positive (e.g., fraud) class versus the probability that it belongs to the negative (e.g., non-fraud) class by a logistic function. Discriminant Analysis Classifier (DAC) assumes that the data from different classes are generated based on different Gaussian distributions. In the training phase, the fitting function calculates the parameters of a Gaussian distribution for each class; while in the testing stage, the trained classifier finds the class with the smallest misclassification cost.

4.4. *Workflow*

The proposed framework includes a training stage as well as a testing stage as shown in Figure 7. The testing dataset is first split into three parts, including a training set, a validation set, and a testing set. In the training phase, the training set conditional probabilities are calculated to build a hierarchical model for all concepts from the training label information.

Next, for all the validation video shots and N concepts, N concept detection models are trained such that for each video shot, the n^{th} model outputs a score measuring the likelihood that concept n exists in that video shot. For this part, all kinds of classifiers can be employed to generate different prediction scores, which may lead to different score-based correlations from the same dataset. That is, for the same C_T (like "cow") in the TRECVID dataset, different hierarchies can be generated based on different classifiers applied. In the aforementioned example, ten score vectors of the positively related concepts and ten score vectors of the negatively related concepts are put together to train an integration model.

In the testing step, each testing video shot of the target concept is plugged into all concept detection models to generate the corresponding testing scores for the related concepts chosen. These scores are then input to the trained score integration model to generate a new set of re-ranked scores. Please note that the scores of the target concept may or may not be used, as shown in Figure 6, depending on whether they are chosen in the training phase or not. Finally, the new output scores are evaluated.

5. Experiments and Results

5.1. *Dataset*

In the experiment, the IACC.1.A and IACC.1.B datasets are chosen from the semantic indexing (SIN) task of the TRECVID 2015 benchmark [53], which aims to detect the semantic concept contained within a video shot. The task assign IACC.1.A as the training dataset and IACC.1.B as the testing dataset. There are several challenges for the SIN task, such as data imbalance, scalability, and the semantic gap [54][55] as mentioned earlier.

The TRECVID conference series encourage research in information retrieval and provide a huge number of videos for training, and there are more than 300 hours in IACC.1.A and IACC.1.B datasets. By extracting keyframes from each video shot, totally 262,911 training data instances are generated. We further divide the IACC.1.B dataset into a validation set with 68,663 data instances and a testing set with the same number of data instances [56].

In this dataset, totally 346 concepts are given, including many popular semantic concepts like "Face", "Vehicle", and "Violent" which are common and appear in many research papers. The list of concepts and the detailed explanations can be found in [27]. In this paper, we download the detection scores from the DVMM Lab of Columbia University [57] for all video shots, who ranked the first several years in the TRECVID competition. The TRECVID 2015 training labels are also utilized to increase the number of ground truth in the negative association selection component. The proposed multimedia big data mining system is tested using some of the results from our previous work as shown in [45][58].

5.2. *Evaluation metrics*

The average precision (AP) value, a widely used metric in the multimedia concept retrieval domain, is used in the evaluation. For a given concept, $Pre(i)$ indicates the precision of the i^{th} data instance in the ranking list. ψ is for the number of the retrieved data instances; while G_n is for the total number of data instances containing that concept in the database. $Min(G_n, \psi)$ indicates the smaller value of G_n and ψ. The average precision at ψ (i.e., $AP@\psi$) is defined in Equation (3). By generating the AP values for all the target rare concepts and calculating the mean value of them, the mean average precision (MAP) value is used to capture the ranking information.

$$AP@\psi = \sum_{i=1}^{\psi} \frac{Pre(i) \times rel(i)}{Min(G_n, \psi)}, \tag{3}$$

$$\text{where } rel(i) = \begin{cases} 1, & \text{if instance } i \text{ is positive;} \\ 0, & \text{otherwise.} \end{cases}$$

5.3. *Experimental results*

Since we target on imbalanced concept retrieval in this paper, 20 most rare concepts with an average P/N ratio of 0.0001 are chosen. Among the video shots in the testing dataset, each of them have no more than 10 video shots in the dataset. These 20 concepts are: “Car Crash”, “Cigar Boats”, “Crustacean”, “High Security Facility”, “Helicopter Hovering”, “Cetacean”, “Military Buildings”, “Rpg”, “Prisoner”, “Police Truck”, “Colin Powell”, “Earthquake”, “Oil Drilling Site”, “Rescue Helicopter”, “Dolphin”, “Security Checkpoint”, “Fire Truck”, “Whale”, “Cows”, and “Yasser Arafat”.

The experimental results are shown in Table 2. The “Baseline” one is calculated using the raw scores directly from the classifiers in [57]. Though the scores here were the best prediction scores, it still performs bad on rare concept retrieval because of the extremely skewed distributions. As mentioned in Section 4.3, we use different classifiers including Support Vector Machine (SVM), Naive Bayes (NB), Random Forest (RF), Logistic Regression (LR), and Discriminant Analysis Classifier (DAC) to re-rank those scores. “LL” is for label-label correlations, which means only using the correlations calculated by the label information in the training dataset. Comparatively, “SL” stands for score-label correlations, which is the main contribution of this paper.

The results clearly show that if the target concepts are extremely rare, using only correlations calculated by the label information from the training dataset does not help and can even downgrade the results. Table 2 shows that we achieve a better score enhancement when using the information generated by the score-label correlations, in comparison with that using the label-label correlations for every classifier.

Table 2. Experimental results.

Framework	MAP10	MAP20	MAP50	MAP100	MAP200	MAP500
Baseline	0.0446	0.0438	0.0312	0.0318	0.0322	0.0302
SVM (LL)	0.0125	0.0125	0.0125	0.0125	0.0130	0.0130
SVM (SL)	0.0167	0.0167	0.0167	0.0088	0.0088	0.0090
NB (LL)	0.0056	0.0142	0.0142	0.0124	0.0124	0.0124
NB (SL)	0.0056	0.0146	0.0146	0.0113	0.0132	0.0137
RF (LL)	0.0375	0.0408	0.0297	0.0215	0.0215	0.0215
RF (SL)	0.0426	0.0460	0.0401	0.0417	0.0417	0.0417
LR (LL)	0.0234	0.0309	0.0335	0.0278	0.0256	0.0230
LR (SL)	0.0467	0.0532	0.0554	0.0551	0.0535	0.0529
DCA (LL)	0.0532	0.0577	0.0554	0.0485	0.0462	0.0436
DCA (SL)	0.1130	**0.1130**	0.0856	0.0733	0.0711	0.0614
Proposed	**0.1321**	0.1101	**0.0916**	**0.0885**	**0.0850**	**0.0681**

Albeit with the imprecise raw scores on rare concepts, the proposed framework can successfully re-rank and enhance the results as can be seen from Table 2.

Since the Naive Bayes (NB) approach is based on applying the Bayes' theorem with strong independence assumptions between the attributes, which is not true in our case (inter-concept correlations), it performs the worst. Furthermore, because of the nature of Random Forrest (RF) (i.e., random tree selected), we run it three times and the results are averaged. Our proposed framework is presented in the "Proposed" column which also includes information from negative correlations. It uses the correlations found in Section 4.2 and integrates the scores from those negative-related concepts.

6. Conclusions and future work

Rare concept retrieval is a challenge task due to the nature of the imbalanced datasets. Since the data instances in the majority class usually overshadows those in the minority class, it is hard to get acceptable retrieval results when the target concept is a rare concept. In this paper, we propose a score re-rank system using the label-score correlations to leverage the semantic concept retrieval task from the video shots. Our experimental results clearly show the effectiveness of the proposed framework and how it can successfully enhance the prediction scores of the rare concepts.

The label-score correlations also work like inference rules which can provide a clue for how to define a rare concept. Suppose we have the data of an unknown kind of "animal", using the proposed framework can help find the relationships between several known animals and concepts with it. Considering the "cow" example, we can now better answer the question of what is a "cow". Similarly, we can somehow define a new concept, a new species, or a new object, even though we do not know what it is now. This kind of definitions is very helpful to the fields of information retrieval and knowledge discovery, which can be further investigated as the future work. Another research direction is to find an efficient way to build larger concept hierarchies. When we have thousands of concepts in a dataset, the trees will be much more complicated and thus more research efforts are needed.

References

[1] Y. Yan and M.-L. Shyu, "Enhancing rare class mining in multimedia big data by concept correlation," in *Proceedings of the 2016 IEEE International Symposium on Multimedia (ISM)*, December 2016, pp. 281–286.

[2] S.-C. Chen, M.-L. Shyu, and R. Kashyap, "Augmented transition network as a semantic model for video data," *International Journal of Networking and Information Systems*, vol. 3, no. 1, pp. 9–25, 2000.

[3] S.-C. Chen, M.-L. Shyu, and C. Zhang, "An intelligent framework for spatio-temporal vehicle tracking," in *Proceedings of the 4th IEEE International Conference on Intelligent Transportation Systems*, August 2001, pp. 213–218.

[4] L. Lin and M.-L. Shyu, "Weighted association rule mining for video semantic detection," *International Journal of Multimedia Data Engineering and Management*, vol. 1, no. 1, pp. 37–54, 2010.

[5] M.-L. Shyu, S.-C. Chen, and R. 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.

[6] S. Sadiq, Y. Yan, M.-L. Shyu, S.-C. Chen, and H. Ishwaran, "Enhancing multimedia imbalanced concept detection using vimp in random forests," in *2016 IEEE 17th International Conference on Information Reuse and Integration (IRI)*, July 2016, pp. 601–608.

[7] S. Pouyanfar and S.-C. Chen, "Semantic event detection using ensemble deep learning," in *The IEEE International Symposium on Multimedia (IEEE ISM)*, CA, USA, 2016, pp. 203–208.

[8] Y. Yan, Y. Liu, M.-L. Shyu, and M. Chen, "Utilizing concept correlations for effective imbalanced data classification," in *Proceedings of the IEEE 15th International Conference on Information Reuse and Integration*, Aug 2014, pp. 561–568.

[9] S.-C. Chen, M.-L. Shyu, and C. Zhang, "Innovative shot boundary detection for video indexing," in *Video Data Management and Information Retrieval*, S. Deb, Ed. Idea Group Publishing, 2005, pp. 217–236.

[10] 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," in *Proceedings of the Third International Workshop on Multimedia Data Mining, in conjunction with the 8th ACM International Conference on Knowledge Discovery & Data Mining*, July 2002, pp. 100–108.

[11] X. Li, S.-C. Chen, M.-L. Shyu, and B. Furht, "Image retrieval by color, texture, and spatial information," in *Proceedings of the 8th International Conference on Distributed Multimedia Systems*, September 2002, pp. 152–159.

[12] L. Lin, G. Ravitz, M.-L. Shyu, and S.-C. Chen, "Video semantic concept discovery using multimodal-based association classification," in *Proceedings of the IEEE International Conference on Multimedia & Expo*, July 2007, pp. 859–862.

[13] ——, "Effective feature space reduction with imbalanced data for semantic concept

detection," in *Proceedings of the IEEE International Conference on Sensor Networks, Ubiquitous, and Trustworthy Computing*, 2008, pp. 262–269.

[14] S.-C. Chen, S. Sista, M.-L. Shyu, and R. Kashyap, "Augmented transition networks as video browsing models for multimedia databases and multimedia information systems," in *Proceedings of the 11th IEEE International Conference on Tools with Artificial Intelligence*, 1999, pp. 175–182.

[15] 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," in *Proceedings of the 7th IEEE International Symposium on Multimedia*, Dec 2005, pp. 37–44.

[16] M. L. Shyu, Z. Xie, M. Chen, and S. C. Chen, "Video semantic event/concept detection using a subspace-based multimedia data mining framework," *IEEE Transactions on Multimedia*, vol. 10, no. 2, pp. 252–259, Feb 2008.

[17] Q. Zhu, L. Lin, M.-L. Shyu, and S.-C. Chen, "Feature selection using correlation and reliability based scoring metric for video semantic detection," in *Proceedings of the Fourth IEEE International Conference on Semantic Computing*, 2010, pp. 462–469.

[18] Y. Yan, M.-L. Shyu, and Q. Zhu, "Supporting semantic concept retrieval with negative correlations in a multimedia big data mining system," *International Journal of Semantic Computing*, vol. 10, no. 02, pp. 247–267, 2016.

[19] Y. Yan, Q. Zhu, M.-L. Shyu, and S.-C. Chen, "A classifier ensemble framework for multimedia big data classification," in *Proceedings of the 2016 IEEE 17th International Conference on Information Reuse and Integration (IRI)*, July 2016, pp. 615–622.

[20] Y. Yan, M. Chen, S. Sadiq, and M.-L. Shyu, "Efficient imbalanced multimedia concept retrieval by deep learning on spark clusters," *International Journal of Multimedia Data Engineering and Management (IJMDEM)*, vol. 8, no. 1, pp. 1–20, 2017.

[21] S.-C. Chen and R. Kashyap, "Temporal and spatial semantic models for multimedia presentations," in *Proceedings of the 1997 International Symposium on Multimedia Information Processing*, 1997, pp. 441–446.

[22] 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, 2001.

[23] X. Li, S.-C. Chen, M.-L. Shyu, and B. Furht, "An effective content-based visual image retrieval system," in *Proceedings of the Computer Software and Applications Conference*, 2002, pp. 914–919.

[24] D. Liu, Y. Yan, M.-L. Shyu, G. Zhao, and M. Chen, "Spatio-temporal analysis for human action detection and recognition in uncontrolled environments," *International Journal of Multimedia Data Engineering and Management*, vol. 6, no. 1, pp. 1–18, Jan. 2015.

[25] M.-L. Shyu, C. Haruechaiyasak, and S.-C. Chen, "Category cluster discovery from distributed www directories," *Information Sciences*, vol. 155, no. 3, pp. 181–197, 2003.

[26] Y. Yan, M. Chen, M.-L. Shyu, and S.-C. Chen, "Deep learning for imbalanced multimedia data classification," in *Proceedings of the 2015 IEEE International Symposium on Multimedia (ISM)*, December 2015, pp. 483–488.

[27] A. F. Smeaton, P. Over, and W. Kraaij, "Evaluation campaigns and TRECVid," in *Proceedings of the 8th ACM International Workshop on Multimedia Information Retrieval*, October 2006, pp. 321–330.

[28] G. E. A. P. A. Batista, R. C. Prati, and M. C. Monard, "A study of the behavior of several methods for balancing machine learning training data," *SIGKDD Explor.*

Newsl., vol. 6, no. 1, pp. 20–29, Jun. 2004.

[29] X. Y. Liu, J. Wu, and Z. H. Zhou, "Exploratory undersampling for class-imbalance learning," *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, vol. 39, no. 2, pp. 539–550, April 2009.

[30] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "Smote: Synthetic minority over-sampling technique," *J. Artif. Int. Res.*, vol. 16, no. 1, pp. 321–357, Jun. 2002.

[31] C. P. Unsworth and G. Coghill, "Excessive noise injection training of neural networks for markerless tracking in obscured and segmented environments," *Neural Comput.*, vol. 18, no. 9, pp. 2122–2145, Sep. 2006.

[32] C. Junfei, W. Qingfeng, and D. Huailin, "An empirical study on ensemble selection for class-imbalance data sets," in *Proceedings of the 5th International Conference on Computer Science Education*, Aug 2010, pp. 477–480.

[33] U. Bhowan, M. Johnston, and M. Zhang, "Developing new fitness functions in genetic programming for classification with unbalanced data," *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, vol. 42, no. 2, pp. 406–421, April 2012.

[34] X. Wang and Q. Ji, "A hierarchical context model for event recognition in surveillance video," in *Proceedings of the 2014 IEEE Conference on Computer Vision and Pattern Recognition*, June 2014, pp. 2561–2568.

[35] B. Ni, Y. Song, and M. Zhao, "Youtubeevent: On large-scale video event classification," in *Proceedings of the 2011 IEEE International Conference on Computer Vision Workshops (ICCV Workshops)*, November 2011, pp. 1516–1523.

[36] Y. T. Yang and C. T. Chiu, "Boosted multi-class object detection with parallel hardware implementation for real-time applications," in *Proceedings of the 2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, May 2014, pp. 7530–7534.

[37] D. T. J. Vreeswijk, C. G. M. Snoek, K. E. A. van de Sande, and A. W. M. Smeulders, "All vehicles are cars: Subclass preferences in container concepts," in *Proceedings of the 2nd ACM International Conference on Multimedia Retrieval.* New York, NY, USA: ACM, 2012, pp. 8:1–8:7.

[38] L. Jiang, S.-I. Yu, D. Meng, Y. Yang, T. Mitamura, and A. G. Hauptmann, "Fast and accurate content-based semantic search in 100m internet videos," in *Proceedings of the 23rd ACM International Conference on Multimedia.* New York, NY, USA: ACM, 2015, pp. 49–58.

[39] T. Meng and M.-L. Shyu, "Concept-concept association information integration and multi-model collaboration for multimedia semantic concept detection," *Information Systems Frontiers*, pp. 1–13, 2013.

[40] S.-C. Chen, A. Ghafoor, and R. L. Kashyap, *Semantic Models for Multimedia Database Searching and Browsing.* Norwell, MA, USA: Kluwer Academic Publishers, 2000.

[41] M.-L. Shyu, S.-C. Chen, M. Chen, and C. Zhang, "A unified framework for image database clustering and content-based retrieval," in *Proceedings of the 2Nd ACM International Workshop on Multimedia Databases*, ser. MMDB '04. New York, NY, USA: ACM, 2004, pp. 19–27.

[42] M.-L. Shyu, S.-C. Chen, M. Chen, C. Zhang, and K. Sarinnapakorn, "Image database retrieval utilizing affinity relationships," in *Proceedings of the 1st ACM International Workshop on Multimedia Databases*, ser. MMDB '03. New York, NY, USA: ACM, 2003, pp. 78–85.

[43] M.-L. Shyu, T. Quirino, Z. Xie, S.-C. Chen, and L. Chang, "Network intrusion detec-

tion through adaptive sub-eigenspace modeling in multiagent systems," *ACM Trans. Auton. Adapt. Syst.*, vol. 2, no. 3, Sep. 2007.

[44] P. Mettes, D. C. Koelma, and C. G. M. Snoek, "The imagenet shuffle: Reorganized pre-training for video event detection," *CoRR*, vol. abs/1602.07119, 2016.

[45] T. Meng, Y. Liu, M.-L. Shyu, Y. Yan, and C.-M. Shu, "Enhancing multimedia semantic concept mining and retrieval by incorporating negative correlations," in *Proceedings of the IEEE International Conference on Semantic Computing*, June 2014, pp. 28–35.

[46] J. A. Suykens and J. Vandewalle, "Least squares support vector machine classifiers," *Neural processing letters*, vol. 9, no. 3, pp. 293–300, 1999.

[47] S.-i. Amari and S. Wu, "Improving support vector machine classifiers by modifying kernel functions," *Neural Networks*, vol. 12, no. 6, pp. 783–789, 1999.

[48] K. P. Murphy, "Naive bayes classifiers," *University of British Columbia*, 2006.

[49] L. Breiman, "Random forests," *Machine Learning*, vol. 45, no. 1, pp. 5–32, 2001.

[50] M. R. Segal, "Machine learning benchmarks and random forest regression," *Center for Bioinformatics & Molecular Biostatistics*, 2004.

[51] L. Lin, G. Ravitz, M.-L. Shyu, and S.-C. Chen, "Effective feature space reduction with imbalanced data for semantic concept detection," in *Proceedings of the IEEE International on Sensor Networks, Ubiquitous, and Trustworthy Computing*, June 2008, pp. 262–269.

[52] T. Meng and M.-L. Shyu, "Leveraging concept association network for multimedia rare concept mining and retrieval," in *Proceedings of the IEEE International Conference on Multimedia and Expo*, Melbourne, Australia, July 2012.

[53] P. Over, G. Awad, M. Michel, J. Fiscus, G. Sanders, W. Kraaij, A. F. Smeaton, G. Quéenot, and R. Ordelman, "Trecvid 2015 – an overview of the goals, tasks, data, evaluation mechanisms and metrics," in *Proceedings of TRECVID 2015.* NIST, USA, 2015.

[54] Q. Zhu, L. Lin, M.-L. Shyu, and D. Liu, "Utilizing context information to enhance content-based image classification," *International Journal of Multimedia Data Engineering and Management*, vol. 2, no. 3, pp. 34–51, 2011.

[55] L. Lin, C. Chen, M.-L. Shyu, and S.-C. Chen, "Weighted subspace filtering and ranking algorithms for video concept retrieval," *IEEE Multimedia*, vol. 18, no. 3, pp. 32–43, March 2011.

[56] A. Karpathy, G. Toderici, S. Shetty, T. Leung, R. Sukthankar, and L. Fei-Fei, "Large-scale video classification with convolutional neural networks," in *Proceedings of the 2014 IEEE Conference on Computer Vision and Pattern Recognition*, ser. CVPR '14. Washington, DC, USA: IEEE Computer Society, 2014, pp. 1725–1732.

[57] Y.-G. Jiang, "Prediction scores on TRECVID 2010 data set," http://www.ee.columbia.edu/ln/dvmm/CU-VIREO374/, 2010, last accessed on September 2011. [Online]. Available: http://www.ee.columbia.edu/ln/dvmm/CU-VIREO374/

[58] Y. Yan, M.-L. Shyu, and Q. Zhu, "Negative correlation discovery for big multimedia data semantic concept mining and retrieval," in *Proceedings of the 2016 IEEE Tenth International Conference on Semantic Computing (ICSC)*, February 2016, pp. 55–62.