

Mining High-Level Features from Video using Associations and Correlations

Lin Lin and Mei-Ling Shyu
Department of Electrical and Computer Engineering
University of Miami, Coral Gables, FL 33124, USA
l.lin2@umiami.edu, shyu@miami.edu

Abstract—Association rule mining (ARM) has been studied in the areas of content-based multimedia retrieval and semantic concept detection due to its high efficiency and accuracy. Two important processes in mining the association rules for classification are rule generation and rule selection. In this paper, a novel high-level feature detection framework using the ARM technique together with the correlations among the feature-value pairs is proposed. A new association rule mining (ARM) algorithm has been developed, where the N -feature-value pairs are generated using a combined measure based on (1) the existence of the $(N-1)$ -feature-value pairs (where N is larger than 1), (2) the correlation between different N -feature-value pairs and the concept classes through Multiple Correspondence Analysis (MCA), and (3) the similarity representing the harmonic mean of the inter-similarity and intra-similarity. The final association classification rules are selected by using the calculated harmonic mean of the similarity values. The proposed framework enables the automatic discovery and generation of the N -feature-value pair association rules from the 1-feature-value pairs for classification. Experimenting with 15 high-level features (concepts) and benchmark data sets from TRECVID, our proposed framework achieves promising performance and outperforms three other well-known classifiers (Decision Trees, Support Vector Machine, and Neural Networks) which are commonly used for performance comparison in the TRECVID community.

Keywords-Association rule mining; Multiple Correspondence Analysis (MCA); Concept detection

I. INTRODUCTION

Association rule mining (ARM) has been used to automatically detect the high-level features (concepts) from the video data, with the attempt to address the challenges such as bridging the semantic gap between the high-level concepts and low-level features. For example, [1] [2] [3] have employed the ARM techniques in exploring knowledge from images. Utilizing ARM in order to generate the association rules from a list of feature-value pairs, two measures are considered in the traditional algorithms: *support* which is the proportion of transactions in the database that contain the feature-value pairs, and *confidence* which is a measure of the accuracy of the rule. A frequently used algorithm is *Apriori* [4]. The process of mining association rules could be summarized to two main steps: finding the frequent itemsets which satisfy the minimum support threshold, and generating rules using the frequent item sets which satisfy the minimum confidence threshold.

In [5], a semantic indexing and event detection framework for basketball videos was proposed. The authors developed the measures including temporal distance, temporal support, and confidence for video associations by using distinct features from image, audio, and caption text. The temporal distance is the temporal identification difference of the neighboring shots that contain these two items, and the upper bound that the temporal distance must comply with is the temporal distance threshold. If the temporal distance threshold is ignored, the temporal support and confidence would be considered as the same as the traditionally ones. However, by using this threshold, the temporal support is defined as the number of times the association appears sequentially and each time this association appears, and the temporal distance between any two neighboring items should satisfy the given temporal distance threshold. The temporal confidence is defined as the ratio between the temporal support and the maximal number of possible occurrences of the association.

Another example applying ARM for multimedia retrieval is presented in [6], in which a multi-layer hierarchical framework for generic sports event analysis in sports video was proposed. The leaf nodes of the hierarchical tree are the extracted events from the video clips. The Apriori algorithm is used to mine the rules, in other words, the associations between these events are used to extract those concepts. The hierarchical approach avoids the shot detection and clustering process that are the necessary steps in most of concept and event detection frameworks. At each level of the tree structure, only simple frame-based features are used to classify the video. Due to the fact that the number of frames to be processed are degraded level by level, the computational cost is reduced remarkably.

In [7], a hierarchical temporal association mining approach was developed to automatically capture the optimal temporal patterns significant for characterizing the interesting events in soccer video. It adaptively determines the essential thresholds which are generally defined manually in the traditional association mining approaches. There are two considerations for the use of patterns. First, an event is characterized by both attribute type and its occurrence frequency, because several close views appear in a temporal window might indicate an interesting event in the soccer video. Second, given a reasonable size of temporal window,

the appearance order of the pattern could be captured, which is important in characterizing a target event. For instance, a goal event could be following a penalty kick. An extended Apriori algorithm was proposed to mine the temporal patterns.

In addition to detect the interesting events in sports video, the ARM technique has also been studied in the feedback systems [8], surveillance systems [9], and concept detection systems [10] [11] [12] [13] in various video types. In [10], a framework discovering the concepts from news TV broadcasts using the traditional ARM technique was proposed. After getting pure positive rules from all the positive instances, the negative instances are simply classified through these rules as fuzzy negative set and pure negative set. It discards all fuzzy negative training instances and uses the pure negative training instances to generate the pure negative rule set using Apriori as the way the positive rules are generated. In [13], the ARM technique was used to discover the relationship between concepts. The prediction value of the detector indicates the likelihood that the detector regards the presence of a certain concept. The association rules are combined with the prediction values as a ranking strategy to improve the detection accuracy on experimenting the TRECVID 2005 data.

In our previous study [11], the *Multiple Correspondence Analysis* (MCA) method was utilized as the rule generation mechanism in associative classification using the 1-feature-value pair rules. The approach assisted the classifiers to detect more positive instances in the testing data set without misclassifying too many negative instances of the investigated concepts while testing on the TRECVID 2007 data. Further, we extended the method in [11] to identify the correlation between the 2-feature-value pairs and concept classes, and use both the 1-feature-value pair rules and 2-feature-value pair rules as the final rule set for classification in [12]. The pairs are generated by using the correlation only and all rules are used for classification since no selecting algorithm is applied to select the rules. Therefore the computational cost is large.

One way to reduce the complexity is to prune the feature-value pairs during generation. A traditional way as in Apriori is to delete the candidate N -feature-value pair if not all combinations of the $(N-1)$ -feature-value pairs in the previous stage exist. For example, at the $N=3$ stage, in order to generate the set $\{A_1, A_2, A_3\}$, all the subset $\{A_1, A_2\}$, $\{A_1, A_3\}$, and $\{A_2, A_3\}$ have to be kept in the $(N-1)=2$ stage. After generating the association rules, rule selection is always applied before classification to improve the performance. The support, confidence, and length of each rule are most commonly used thresholds. For instance in [14], the strategy is the longer the rules, the higher the priority of the rules. If two rules have the same length, the rules are sorted according to their confidence values. If two rules have the same confidence value, then the rules are ranked

by their support values. A combined measure to select rules was introduced in [15]. The authors proposed a personalized association rule ranking method based on semantic similarity between the rules, the keywords assigned by user's interests, and the statistical information like support, confidence, and chi-square value.

In this paper, a novel high-level feature detection framework is proposed which is facilitated with a new association rule mining (ARM) algorithm that considers a combined measure for feature-value pair rule generation and a similarity measure for association rule selection. Our proposed framework consists of three main stages, namely (i) the preparation of the training data set and testing data set, which includes low-level feature extraction, feature normalization, data splitting, and discretization; (ii) the generation of the N -feature-value pairs based on the $(N-1)$ -feature-value pairs (when N is greater than 1), the correlation between the extracted low-level features and the concept classes, the similarity calculation, and rule evaluation using the training data set; and (iii) the selection of the rules and the final high-level feature detection (classification).

To evaluate our proposed framework, the high-level features from TRECVID 2007 and 2008 [16] are used to compare the performance between our proposed framework and the well-known *decision tree* classifier, *support vector machine* classifier, and *neural network* classifier. Overall, our proposed framework outperforms all the three classifiers. Furthermore, it is important to mention that the proposed framework significantly outperforms the support vector machine classifier, which is the most recommended classifier in the research community for performance evaluation using the TRECVID data sets.

This paper is organized as follows. The proposed framework and detailed discussion on each component are presented in Section II. Section III discusses on the experiments as well as the result analysis. In Section IV, the conclusions of paper are given.

II. THE PROPOSED HIGH-LEVEL FEATURE DETECTION FRAMEWORK

This paper proposes a novel framework that performs high-level feature (concept) detection from video using the ARM technique together with the correlations among the feature-value pairs is proposed. In the proposed framework, a combined measure is used for association rule generation and a rule selection strategy is performed to optimize the final rule set for classification.

Our proposed framework is shown in Figure 1 which consists of three main stages. These three stages are separated in three boxes enclosed by the dash lines in Figure 1.

- 1) The first stage is data preparation. It includes (a) low-level feature extraction, (b) feature normalization, (c) data splitting, and (d) discretization.

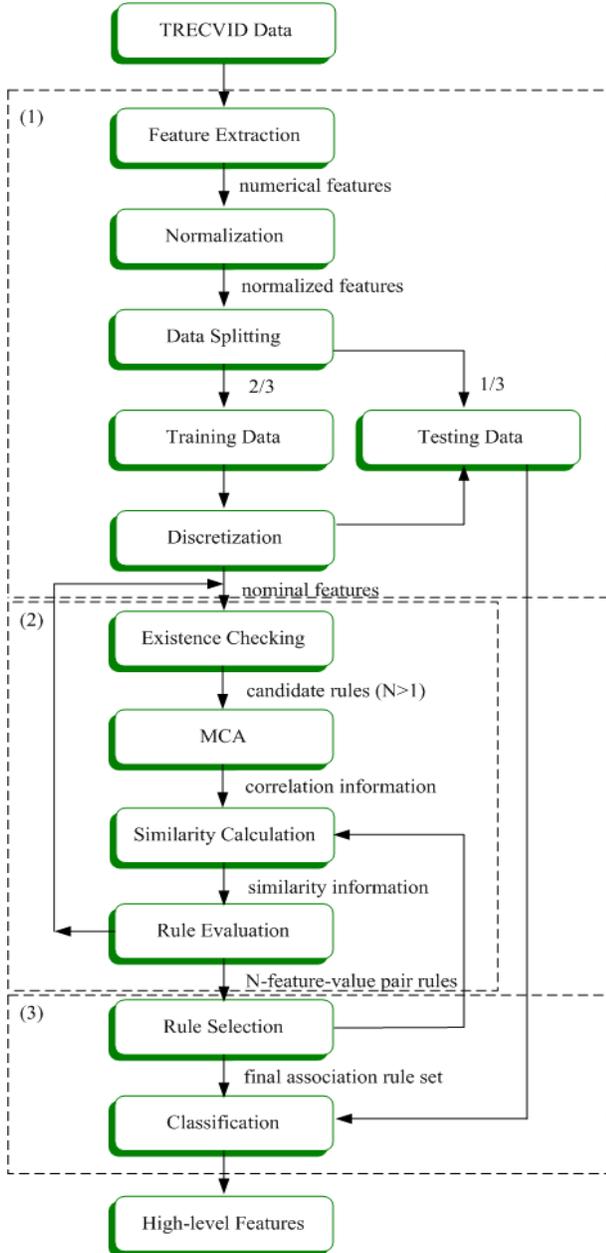


Figure 1. The Proposed Framework

- 2) The second stage is N -feature-value pair rules generation. It works based on (a) checking the existence of the $(N-1)$ -feature-value pairs (where N is larger than 1) and getting the candidate N -feature-value pairs, (b) the correlation between the extracted low-level features and the concept classes through Multiple Correspondence Analysis (MCA) to identify the N -feature-value pairs as the rules that better represent each one of the investigated concepts, (c) the similarity of rules representing the harmonic mean of inter-similarity and intra-similarity of rules, (d) rule

evaluation through the training data set to get the threshold values for parameters automatically.

- 3) The third stage is rule selection and classification. Rule selection is processed in the following way that all the positive rules are stored and then the negative rules are selected by using the mean of the similarity of all negative rules. Then classification is done by using the final association rule set, where the concept class is determined by the majority class of the matched association rules.

A. The First Stage: Data Preparation

Since the shot boundary information has been provided by [16], the video shot boundary detection is beyond the scope of this paper, and the audio and visual features extracted in this paper are shot-based. Totally, 28 continuous numerical features that were used in [10] are extracted. The feature values are normalized to lie in the $[0, 1]$ range, as introduced in [17], simply by subtracting the minimum value and dividing by the difference between the maximum and the minimum values. Then the data set is split into two parts, two-third of the data is used for training and one-third of the data is used for testing, and the cross-validation method ensures that each data instance in the data set would be tested at least once as in [12].

Due to the fact that the ARM requires the input data to be nominal, all the extracted features are discretized using the method discussed in [18]. The training data set is discretized into bins (partitions) which are then used to discretize the testing data set. If there is only one resulting partition by using the aforementioned method, another discretization process is executed by using the average values of the feature for the disparity measure to construct two partitions for that feature. Please note that all the partitions generated by the discretization process are called the *feature-value pairs* in our study.

B. The Second Stage: Rule Generation

After the features are discretized, the training data are ready for rule generation. As demonstrated in Figure 1, there are four steps at the rule generation stage. The method to generate the 2-feature-value pairs was discussed in [12], that after the 1-feature-value pairs have been generated, the 2-feature-value pairs are generated by pairing the 1-feature-value pairs according to whether both of the 1-feature-value pair rules were generated for the same class. However, this is not applicable to the N -feature-value pairs due to the large computational cost. Therefore, a pruning strategy is developed. That is, the possible candidate N -feature-value pairs are first listed, and then simply checking the existence of the $(N-1)$ -feature-value pairs (where N is larger than 1) using the method introduced in Section I, the candidate N -feature-value pairs are pruned.

Next, MCA is utilized to calculate the correlation between the nominal feature-value pairs and the concept classes, and the N -feature-value pairs as the rules that better represent each target concept are discovered. For each rule, the inter-similarity is defined as the similarity between the rule and the data instances belonging to different classes; while the intra-similarity is the similarity between the rule and the data instances belonging to the same class. The harmonic mean of the inter-similarity and intra-similarity values are calculated as a further step to generate the N -feature-value pair rules. The threshold values for correlation and similarity information are gained by evaluating the candidate rules through the training data set. The details description is given in the Section Section II-D

C. The Third Stage: Rule Selection and Classification

Since the detection of the target concept is the main concern, the positive rules are more important. Moreover, for some concepts, there are much more negative rules than the positive rules. Therefore, the strategy is that all the association rules (positive rules) indicating the concept class are kept. Then by using the mean of the similarity values of all rules for the non-concept class (negative rules), the negative rules are ranked. Finally, the selected N -feature-value pair rules are used as the association rules for mining the high-level features (classification).

In [11] [12], the utilization of MCA to analyze the data instances described by a set of low-level features and high-level concepts has been explored. MCA is an extension of the standard correspondence analysis to more than two variables [19]. Therefore, MCA can be extended to capture the correlation between the N -feature-value pairs and the target concept (class), and we are able to project the multimedia data into a new space using the first and the second principle components. The inner product of all possible N -feature-value pairs and the classes can be calculated and the angles between the N -feature-value pairs and the classes are used as a measurement to represent their correlations. The main idea of utilizing MCA is that the smaller the angle is, the higher the correlation is. Therefore, each feature-value pair will be assigned to one class (positive or negative) only, and the candidate N -feature-value pair rules are available for the classification/detection of the particular target concept.

D. The Details on Rule Generation and Rule Selection

The pseudo-code for generating the 1-feature-value pair rules is presented as follows. Assume that all the 1-feature-value pairs gained from discretization are considered as the candidate pairs ($pair_s^{(1)}$), and $S^{(1)}$ is the total number of candidates. By applying MCA to calculate the angles between the pairs in $pair_s^{(1)}$ and the classes $class_r$ (either $class_{pos}$ or $class_{neg}$), the candidate 1-feature-value pair rules are available. Then the similarity of these association rules could be obtained. The positive 1-feature-value pair

rules whose angles are smaller than a certain threshold ($threshold_1^{(1)}$) and their similarity values are larger than another threshold ($threshold_3^{(1)}$) are selected. Similarly, the negative 1-feature-value pair rules whose angles are smaller than the threshold ($threshold_2^{(1)}$) and their similarity values are larger than the threshold ($threshold_4^{(1)}$) are selected. Here, i is the index of the rules.

RULE-GENERATION (N=1)

```

1   $i \leftarrow 1$ ;
2   $n \leftarrow 1$ ;
3  for  $s \leftarrow 1$  to  $S^{(1)}$ 
4    compute  $angle_s^{(1)}$  by applying MCA;
5     $rule_s \leftarrow pair_s^{(1)} \Rightarrow class_r$ ;
6    compute  $similarity_s^{(1)}$  of  $rule_s$ ;
7    if  $angle_s^{(1)} < threshold_1^{(1)}$  and
8    if  $similarity_s^{(1)} > threshold_3^{(1)}$  then
9       $rule_i \leftarrow pair_s^{(1)} \Rightarrow class_{pos}$ ;
10      $i \leftarrow i + 1$ ;
11   if  $angle_s^{(1)} < threshold_2^{(1)}$  and
12   if  $similarity_s^{(1)} > threshold_4^{(1)}$  then
13      $rule_i \leftarrow pair_s^{(1)} \Rightarrow class_{neg}$ ;
14      $i \leftarrow i + 1$ ;

```

RULE-GENERATION FOR POSITIVE RULES (N>1)

```

1   $n \leftarrow 2$ ;
2  while (1)
3    if  $class_r = class_{pos}$ 
4      for  $s \leftarrow 1$  to  $S_{pos}^{(n)}$ 
5        check  $pair_s^{(n)}$ ;
6        if  $pair_s^{(n)} \neq NULL$  then
7          compute  $angle_s^{(n)}$  by applying MCA;
8           $rule_s \leftarrow pair_s^{(n)} \Rightarrow class_{pos}$ ;
9          compute  $similarity_s^{(n)}$  of  $rule_s$ ;
10         if  $angle_s^{(n)} < threshold_1^{(n)}$  and
11         if  $similarity_s^{(n)} > threshold_3^{(n)}$  then
12            $rule_i \leftarrow pair_s^{(n)} \Rightarrow class_{pos}$ ;
13            $i \leftarrow i + 1$ ;
14         if  $pair_s^{(n+1)}$  is possible then
15            $n \leftarrow n + 1$ ;
16         if not possible then
17           break.

```

When n is greater than 1, a combination of those $(n-1)$ -feature-value pairs that do not belong to the same feature but indicate to the same concept (class) is considered as a candidate n -feature-value pair, and $S^{(n)}$ is the total number of candidates which equals the sum of the number of the candidates for the positive class $S_{pos}^{(n)}$ and the number of the candidates for the negative class $S_{neg}^{(n)}$. If all the combinations are considered, the computational cost is large.

Therefore, when n is greater than 1, the $pair_s^{(n)}$ is checked first to reduce the computational cost.

RULE-GENERATION FOR NEGATIVE RULES ($N > 1$)

```

1   $n \leftarrow 2$ ;
2  while (1)
3    if  $class_r = class_{neg}$ 
4      for  $s \leftarrow 1$  to  $S_{neg}^{(n)}$ 
5        check  $pair_s^{(n)}$ ;
6        if  $pair_s^{(n)} \neq NULL$  then
7          compute  $angle_s^{(n)}$  by applying MCA;
8           $rule_s \leftarrow pair_s^{(n)} \Rightarrow class_{neg}$ ;
9          compute  $similarity_s^{(n)}$  of  $rule_s$ ;
10         if  $angle_s^{(n)} < threshold_2^{(n)}$  and
11         if  $similarity_s^{(n)} > threshold_4^{(n)}$  then
12            $rule_i \leftarrow pair_s^{(n)} \Rightarrow class_{neg}$ ;
13            $i \leftarrow i + 1$ ;
14         if  $pair_s^{(n+1)}$  is possible then
15            $n \leftarrow n + 1$ ;
16         if not possible then
17           break.

```

The idea is that all the subsets containing the $(n-1)$ -feature-value pairs with the contribution on combining the n -feature-value-pair are kept at the previous step. Then for the candidate n -feature-value pairs, MCA is applied to calculate the angles and the similarity values are calculated to generate the n -feature-value pair rules for $class_r$. The process will continue until no new candidate pairs are available. The pseudo-code for generating both the positive N -feature-value pair rules and the negative N -feature-value pair rules, when N is larger than 1, are presented.

After the N -length candidate rules have been generated, the inter-similarity and intra-similarity values are calculated to evaluate each rule. For a certain concept C_j , assume that there are T_{jp} positive data instances, T_{jn} negative data instances, R_{kp} N -feature-value pairs association rules generated for C_j , and R_{kn} N -feature-value pairs association rules for the non- C_j class. Here, $ct1$ is the counter of the event that all N feature-value pairs are matched, $ct2$ is the counter when only some of N feature-value pair are matched, and $ct3$ is the counter that none of the feature-value pair is matched. The intra-similarity and inter-similarity values of the rules are defined as follows. Please note that for the similarity values in our definitions, a larger inter-similarity value means that the rule is better, and a larger intra-similarity value also indicates the rule is better.

$$SIntra_{R_{kp}} = \left(\sum_{ct1} 1 + \sum_{ct2} 0.25 + \sum_{ct3} 0 \right) / T_{jp};$$

$$SInter_{R_{kp}} = \left(\sum_{ct1} 0 + \sum_{ct2} 0.75 + \sum_{ct3} 1 \right) / T_{jn};$$

$$SIntra_{R_{kn}} = \left(\sum_{ct1} 1 + \sum_{ct2} 0.25 + \sum_{ct3} 0 \right) / T_{jn};$$

$$SInter_{R_{kn}} = \left(\sum_{ct1} 0 + \sum_{ct2} 0.75 + \sum_{ct3} 1 \right) / T_{jp}.$$

$$I_{kp} = \frac{2 \times SInter_{R_{kp}} \times SIntra_{R_{kp}}}{(SInter_{R_{kp}} + SIntra_{R_{kp}})};$$

$$I_{kn} = \frac{2 \times SInter_{R_{kn}} \times SIntra_{R_{kn}}}{(SInter_{R_{kn}} + SIntra_{R_{kn}})}.$$

Let $I_{kp} \in [0, 1]$ and $I_{kn} \in [0, 1]$, then the harmonic mean of the inter-similarity and intra-similarity values is calculated. The harmonic mean is adopted since the harmonic mean of a list of numbers tends strongly toward the least elements of the list when compared to the arithmetic mean (simple average). If the inter-similarity and intra-similarity values are large, the harmonic mean tends to mitigate the impact of the large values. If the inter-similarity or intra-similarity value is small, the harmonic mean tends to aggravate the impact of the small values.

At this stage, the correlation information (the angle values) and the similarity information (the intra-similarity and inter-similarity values) for each N -feature-value pair rule have been obtained. Then all these candidate rules are evaluated through the training data set. As indicated earlier, a few threshold values need to be pre-defined in our proposed framework. The angle threshold is the one that achieves the highest accuracy when applying the generated rule set with different thresholds to evaluate the training data set. Similarly, applying the generated rule set using different thresholds, the similarity threshold is the one which yields the highest accuracy. The candidate rules corresponding to the angle and similarity values that are larger than the threshold value are selected. From the experiments, the length of the rule N varies from 2 to 5. Please note that in order to reduce the computational cost, the same threshold was used for the positive and negative association rules in [12]. However, in this paper, we have successfully reduced the cost by pruning the feature-value pairs. Therefore, the thresholds of the angle ($threshold_1^{(n)}$ and $threshold_2^{(n)}$) and the thresholds of similarity ($threshold_3^{(n)}$ and $threshold_4^{(n)}$) for positive and negative rules are different and adaptively determined in this framework at each step of N .

The classification process is carried out as follows. For each target concept, each testing data instance in the testing data set will go through all the selected rules (i.e., N -feature-value pairs) so that the matched rules can be identified. The matched rules are those N -feature-value pairs that are found in the testing data instance. The classification for each testing data instance is determined by the majority class label of those matched N -feature-value pairs for the corresponding testing data instance. If there is a tie, meaning that there are equal numbers of matched positive rules and the matched negative rules, then the positive rules are more important

and thus considered as having a higher priority in the high-level feature detection (classification) process.

The same classification procedure is applied to all the concepts, in the manner that for each concept, the N -feature-value pair rules are generated and the selected association rules are used as the classification rules.

III. EXPERIMENTS AND RESULTS

Our proposed framework is validated using the video data available for the TRECVID high-level feature extraction task participants in 2007 and 2008. The high-level features with sufficient amounts of positive data instances to build useful training and testing data sets in our database are *two-people*, *outdoor*, *building*, *vegetation*, *street*, *road*, *sky*, *hand*, *urban*, *waterscape*, *crowd*, *face*, *person*, *animal*, and *walking*, whose descriptions can be found in [16].

The performance evaluation was conducted by comparing the proposed framework with the Decision Tree (C4.5 algorithm), Support Vector Machine (Sequential Minimal Optimization (SMO) algorithm), and Neural Network (Multilayer Perception algorithm) classifiers available in WEKA [17]. Please note that the parameters of these classifiers are set to the default values that WEKA has. The same evaluation metrics (namely, the precision, recall, and F1-score measures as given in Equations (1), (2), and (3)) and the same data sets are used in the comparison, where TP , FP , and FN represent true positive, false positive, and false negative respectively. The 3-fold cross-validation approach is adopted to ensure that each data instance is tested in the experiments.

$$Recall = \frac{TP}{TP + FN}; \quad (1)$$

$$Precision = \frac{TP}{TP + FP}; \quad (2)$$

$$F1score = \frac{2 \times TP}{(TP + FN) + (TP + FP)}. \quad (3)$$

The average precision (Pre), recall (Rec), and F1-score (F1) values from the three folds are presented in Table I. In these tables, the performance obtained from WEKA's Decision Tree (DT), Support Vector Machine (SVM), and Neural Network (NN) are given in columns 2 to 4 respectively; while the performance of our proposed framework (ARM) is shown in the last column.

From Table I, it can be clearly seen that comparing to the other three classifiers, the proposed framework that extracts high-level features from the video data using the ARM technique together with the correlations among the feature-value pairs achieves the promising results. Our proposed framework not only outperforms the DT, SVM and NN classifiers in the average F1-score which considers both recall and precision, but also obtains higher recall values

Table I
PERFORMANCE EVALUATION FOR FIFTEEN HIGH-LEVEL CONCEPTS

Two-people	DT	SVM	NN	ARM
Pre	0.51	0.00	0.44	0.37
Rec	0.19	0.00	0.25	0.80
F1	0.27	0.00	0.32	0.51
Outdoor	DT	SVM	NN	ARM
Pre	0.59	0.59	0.53	0.45
Rec	0.39	0.37	0.48	0.77
F1	0.47	0.46	0.50	0.57
Building	DT	SVM	NN	ARM
Pre	0.57	0.55	0.50	0.42
Rec	0.34	0.27	0.42	0.83
F1	0.42	0.36	0.46	0.56
Vegetation	DT	SVM	NN	ARM
Pre	0.51	0.54	0.45	0.39
Rec	0.35	0.14	0.46	0.86
F1	0.42	0.21	0.45	0.53
Street	DT	SVM	NN	ARM
Pre	0.55	0.58	0.49	0.47
Rec	0.50	0.47	0.49	0.80
F1	0.52	0.52	0.49	0.59
Road	DT	SVM	NN	ARM
Pre	0.58	0.59	0.50	0.45
Rec	0.34	0.35	0.47	0.81
F1	0.42	0.44	0.48	0.57
Sky	DT	SVM	NN	ARM
Pre	0.62	0.64	0.56	0.48
Rec	0.48	0.45	0.51	0.78
F1	0.54	0.52	0.54	0.59
Hand	DT	SVM	NN	ARM
Pre	0.46	0.33	0.42	0.39
Rec	0.31	0.06	0.40	0.81
F1	0.37	0.10	0.41	0.53
Urban	DT	SVM	NN	ARM
Pre	0.51	0.53	0.47	0.45
Rec	0.41	0.25	0.45	0.76
F1	0.46	0.34	0.46	0.56
Waterscape	DT	SVM	NN	ARM
Pre	0.58	0.65	0.53	0.48
Rec	0.49	0.47	0.52	0.74
F1	0.53	0.54	0.52	0.58
Crowd	DT	SVM	NN	ARM
Pre	0.54	0.78	0.50	0.39
Rec	0.32	0.03	0.49	0.87
F1	0.40	0.06	0.50	0.54
Face	DT	SVM	NN	ARM
Pre	0.60	0.68	0.55	0.47
Rec	0.43	0.39	0.48	0.67
F1	0.50	0.50	0.51	0.55
Person	DT	SVM	NN	ARM
Pre	0.54	0.61	0.48	0.43
Rec	0.39	0.32	0.45	0.77
F1	0.45	0.42	0.47	0.55
Animal	DT	SVM	NN	ARM
Pre	0.62	0.66	0.57	0.48
Rec	0.49	0.50	0.54	0.79
F1	0.54	0.57	0.55	0.59
Walking	DT	SVM	NN	ARM
Pre	0.56	0.59	0.53	0.51
Rec	0.52	0.57	0.52	0.82
F1	0.54	0.57	0.52	0.63

than the DT, SVM and NN classifiers when considering the accuracy of high-level feature detection. The recall values for positive class have been improved because we are able to identify more good association rules for positive class. Although there is a little bit trade-off between precision and recall, the overall F1-score have been improved. These observations demonstrate that our proposed framework using the novel ARM algorithm based on both associations and correlations assists in the classifier to detect more positive data instances labeled with the target high-level features in the testing data set, and improves the classification accuracy of the selected association rules.

IV. CONCLUSION

In this paper, a novel high-level feature mining/detection framework utilizing the associations and correlations among the feature-value pairs and the target concepts is proposed. A new association rule mining (ARM) algorithm has been developed due to its high efficiency and accuracy in the areas of multimedia retrieval and high-level feature (concept/event) detection. In our proposed new ARM algorithm, the N -feature-value pair rules are generated based on a combined measure from (1) the existence of the $(N-1)$ -feature-value pairs (where N is larger than 1), (2) the correlation between different N -feature-value pairs and the high-level features (concept classes) through Multiple Correspondence Analysis (MCA), (3) the similarity value representing the harmonic mean of inter-similarity and intra-similarity, and (4) the evaluation of the rules using the training data to automatically get various thresholds. Then a rule selection strategy is performed to optimize the final rule set. The target concept class is determined by the majority class of the matched selected association rules. The video data from the TRECVID 2007 and 2008 video corpus is used to validate the detection performance of our proposed framework. The experimental results show that our proposed framework demonstrates promising results with better overall recall and F1-score performance over the DT, SVM, and NN classifiers that are commonly used for performance comparison in the TRECVID community.

ACKNOWLEDGMENT

For Mei-Ling Shyu, this research was carried out in part under the auspices of the Cooperative Institute for Marine and Atmospheric Studies (CIMAS), a Joint Institute of the University of Miami and the National Oceanic and Atmospheric Administration, cooperative agreement #NA17RJ1226. "The findings and conclusions in this report are those of the author(s) and do not necessarily represent the views of the funding agency."

REFERENCES

- [1] A. Kouomou-Choupo, L. Berti-Equille, and A. Morin, "Multimedia indexing and retrieval with features association rules mining [image databases]," in *IEEE International Conference on Multimedia and Expo (ICME04)*, June 2004, pp. 1299–1302.
- [2] V. S. Tseng, M.-H. Wang, and J.-H. Su, "A new method for image classification by using multilevel association rules," in *IEEE International Conference on Data Engineering (ICDE05)*, April 2005, pp. 1180–1188.
- [3] H. H. Malik and J. R. Kender, "Clustering web images using association rules, interestingness measures, and hypergraph partitions," in *ACM International Conference on Web Engineering (ICWE06)*, July 2006, pp. 48–55.
- [4] R. Agrawal and R. Srikant, "Fast algorithms for mining association rules in large databases," in *International Conference on Very Large Data Bases (VLDB94)*, September 1994, pp. 487–499.
- [5] X. Zhu, X. Wu, A. K. Elmagarmid, Z. Feng, and L. Wu, "Video data mining: Semantic indexing and event detection from the association perspective," *IEEE Transactions on Knowledge and Data Engineering*, vol. 17, no. 5, pp. 665–677, May 2005.
- [6] M. H. Kolekar and S. Sengupta, "Semantic concept extraction from sports video for highlight generation," in *ACM International Conference on Mobile Multimedia Communications (MobiMedia06)*, September 2006, pp. 1–6.
- [7] M. Chen, S.-C. Chen, and M.-L. Shyu, "Hierarchical temporal association mining for video event detection in video databases," in *IEEE International Workshop on Multimedia Databases and Data Management (MDDM07)*, April 2007, pp. 137–145.
- [8] N. Zhao, S.-C. Chen, and S. H. Rubin, "Automated multimedia systems training using association rule mining," in *IEEE International Conference on Information Reuse and Integration (IRI07)*, August 2007, pp. 373–378.
- [9] F. Anwar and A. Naftel, "Video event modelling and association rule mining in multimedia surveillance systems," in *International Conference on Visual Information Engineering (VIE08)*, August 2008, pp. 426–431.
- [10] L. Lin, G. Ravitz, M.-L. Shyu, and S.-C. Chen, "Video semantic concept discovery using multimodal-based association classification," in *IEEE International Conference on Multimedia and Expo (ICME07)*, July 2007, pp. 859–862.
- [11] —, "Correlation-based video semantic concept detection using multiple correspondence analysis," in *IEEE International Symposium on Multimedia (ISM08)*, December 2008, pp. 316–321.
- [12] L. Lin, M.-L. Shyu, G. Ravitz, and S.-C. Chen, "Video semantic concept detection via associative classification," in *IEEE International Conference on Multimedia and Expo (ICME09)*, July 2009, pp. 418–421.

- [13] K.-H. Liu, M.-F. Weng, C.-Y. Tseng, Y.-Y. Chuang, and M.-S. Chen, "Association and temporal rule mining for post-filtering of semantic concept detection in video," *IEEE Transactions on Multimedia*, vol. 10, pp. 240–251, February 2008.
- [14] Q. Song, M. Shepperd, M. Cartwright, and C. Mair, "Software defect association mining and defect correction effort prediction," *IEEE Transactions on Software Engineering*, vol. 32, pp. 69–82, February 2006.
- [15] G. Yang, K. Shimada, S. Mabu, and K. Hirasawa, "A personalized association rule ranking method based on semantic similarity and evolutionary computation," in *IEEE Congress on Evolutionary Computation (CEC08)*, June 2008, pp. 487–494.
- [16] A. F. Smeaton, P. Over, and W. Kraaij, "Evaluation campaigns and TRECVID," in *ACM International Workshop on Multimedia Information Retrieval (MIR06)*, October 2006, pp. 321–330.
- [17] I. H. Witten and E. Frank, *Data Mining: Practical Machine Learning Tools and Techniques*, 2nd ed. Morgan Kaufmann, June 2005.
- [18] U. M. Fayyad and K. B. Irani, "On the handling of continuous-valued attributes in decision tree generation," *Machine Learning*, vol. 8, pp. 87–102, January 1992.
- [19] N. J. Salkind, Ed., *Encyclopedia of Measurement and Statistics*. SAGE Publications, Inc, 2007.