VIDEO SEMANTIC CONCEPT DETECTION VIA ASSOCIATIVE CLASSIFICATION

Lin Lin, Mei-Ling Shyu, Guy Ravitz

Department of Electrical and
Computer Engineering
University of Miami
Coral Gables, FL 33124, USA
l.lin2@umiami.edu, {shyu,ravitz}@miami.edu

Shu-Ching Chen

School of Computing and
Information Sciences
Florida International University
Miami, FL 33199, USA
chens@cs.fiu.edu

ABSTRACT

Associative classification (AC) has been studied in the areas of content-based multimedia retrieval and semantic concept detection due to its high accuracy. The traditional AC algorithm discovers the association rules with the frequency count (minimum support) and ranking threshold (minimum confidence) while restricted to the concepts (class labels). In this paper, we propose a novel framework with a new associative classification algorithm which generates the classification rules based on the correlation between different feature-value pairs and the concept classes by using *Multiple Correspondence Analysis* (MCA). Experimenting with the high-level features and benchmark data sets from TRECVID, our proposed algorithm achieves promising performance and outperforms three well-known classifiers which are commonly used for performance comparison in the TRECVID community.

Index Terms— Associative Classification, Multiple Correspondence Analysis, Concept Detection.

1. INTRODUCTION

The ability to automatically detect high-level semantic concepts from multimedia databases and to narrow the gap between such concepts and the low-level features has been investigated by many researchers. The motivations to extract objects, concepts, or events from images and videos lie in the fact that it could immensely improve the performance of applications such as search engines, video/audio recommenders, and video summarizers.

Classification using *Association Rule Mining* [1] takes advantages of its high accuracy and ability to handle large databases [2]. However, this comes with some challenges [3] such as rule ranking and rule overlapping. One new technique for concept detection, called Association Rule Classification (ARC) or Associative Classification (AC), has been utilized in content-based multimedia retrieval and concept/event detection in recent years. In [4], an image classification approach was presented by using multiple-level association rules based on objects in the images. In our previous work [5], we proposed a framework that discovered three semantic concepts from news TV broadcast using traditional associative classification. Initially, low-level audio-visual features were extracted from the broadcast. Following that, pure positive and pure negative association rules were generated for each concept. Finally, a classifier with a different classification rule ranking strategy was developed for each concept.

One of the first associative classification algorithms introduced in the literature was Classification Based on Associations (CBA) [6]. It generated the association rules with certain support and confidence values based on the Apriori algorithm. Since then, three main research aspects for associative classification have emerged. One is to improve the support and confidence measurements themselves [7][8]. Another one is to use other evaluation criteria such as lift, coverage, leverage, and conviction [1]. The last one is to use an integrated algorithm to generate association rules. For example, in [9], the Classification based on Predictive Association Rules (CPAR) combined the advantages of both associative classification (with high accuracy) and traditional rule-based classifiers named First Order Inductive Learner (FOIL) with less computational cost. CPAR adopted a greedy algorithm to generate more rules directly from the training data, used an expected accuracy to evaluate each rule, and the best k-rules were used in prediction. In our previous study [10], we have utilized the *Multiple Correspondence Analysis (MCA)* methodology as the rule generation mechanism in associative classification. MCA is an extension of standard correspondence analysis to more than two variables [11]. Considering a multimedia database, the columns represent the features and classes, and the rows represent the instances. Using MCA, the correspondence between the features and classes could be captured. Although the classification rules were based on only 1-feature-value pairs in [10], the approach was able to help the classifiers to detect more positive instances in the testing data set without misclassifying too many negative instances of the investigated concepts, especially for the imbalanced data sets.

In this paper, we propose a novel framework which extends MCA to identify the correlation between 2-feature-value pairs and concepts, and use both 1-feature-value pair rules and 2-feature-value pair rules as the rule set for associative classification. By using those rules captured by MCA, we perform classification (concept detection). To evaluate our proposed framework, we used the concepts and data from TRECVID 2007 and 2008 [12], and compared the performance of our proposed framework to that of the well-known *decision tree* classifier (DT), *support vector machine* classifier (SVM), and traditional *association rule* classifier (ARC). The reasons we compared with these three classifiers are (i) DT is the traditional rule-based classifier and the most frequently used one in event/concept detection; (ii) SVM is the most popular classifier for comparison using TRECVID datasets; and (iii) ARC is our initial motivation of improving associative classification. Overall, our proposed framework performs 10% or higher on accuracy evaluation (F1-score) than the three compared classifiers for nine concepts.

This paper is organized as follows. In Section 2, the proposed framework and detailed discussions on its different components are presented. Section 3 discusses our experiments as well as our observations. This paper is concluded in Section 4.

2. THE PROPOSED VIDEO SEMANTIC CONCEPT DISCOVERY FRAMEWORK

We propose a novel video semantic concept detection framework with the following steps. First, a set of 28 low-level audiovisual features is extracted from the video data and normalized. Second, the data is split into training and testing data sets and discretized to generate the association rules. Next, the feature-value pairs from the discretization stage are evaluated and MCA is utilized to get the correlation between feature-value pairs and classes. Finally, classification using the generated rules measured by the correlation is performed.

2.1. Framework Architecture

Our proposed framework is shown in Figure 1. After extracting the audio-visual features (numerical features), each feature is normalized for each video. Next, the whole data set is split into the training data set (two thirds of the data) and the testing data set (one third of the data). In order to implement a 3-fold cross-validation experiment, the positive instances (instances with the target concept label) and negative instances (instances with non-concept labels) are split into three equal parts separately. Two parts of the positive instances and two parts of the negative instances are used as the training set, and the remaining part of the positive instances and the remaining part of the negative instances are used for testing. By doing this, we could ensure that each instance in the data set would be tested at least once.

Due to the facts that associative classification requires the input data to be nominal and all our features are numerical, the features need to be discretized in the next stage. First, the training data was discretized and then the same partitions are used for discretizing the testing data. Next, MCA is applied to calculate the angle between each candidate feature-value pair and each class for the training data. Those 1-feature-value pairs whose corresponding angles are smaller than the first threshold value will be kept. Those selected 1-feature-value pairs are then used to generate the candidate 2-feature-value pairs by pairing every two 1-feature-value pairs that are from different numerical features. These candidate 2-feature-value pairs will apply MCA and those whose corresponding angles are smaller than the second threshold value would be kept. Both threshold values are determined by using the training data. Finally, classification is performed using the associative rules generated by the selected 1-feature-value and 2-feature-value pairs.

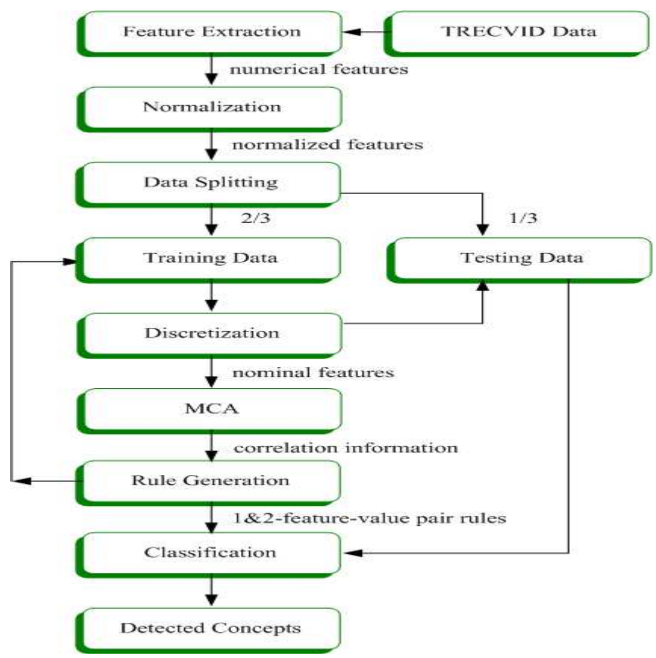

Fig. 1. The proposed framework.

2.2. Associative Classification

An association rule is a relation among items (feature-value pairs in this study) [1]. Let D be a multimedia database consisting of N instances/rows (D_n where n=1 to N), F be a feature set consisting of M numerical features (A_m where m=1 to M), and C be a class set consisting of J discrete classes/concepts (C_j where j=1 to J). The data instances in D are characterized by $M+1$ low-level features/columns (i.e., M numerical values and 1 nominal class label C_j).

After discretization which converts each numerical feature into discrete bins, each feature A_m has K_m nominal feature-value pairs (A_m^i where i=1 to K_m), and the total number of feature-value pairs (i.e., the sum of all K_m) is equal to K. Then each D_n can be described as the feature-value pairs A_m^i and the class C_j. For instance, A_{17} is the feature of the pixel changes, which is converted to 3 bins (i.e., $K_{17} = 3$), and A_{17}^1, A_{17}^2, and A_{17}^3 represent the feature value ranges $[0, 0.32865]$, $(0.32865, 0.5044]$, and $(0.5044, 1]$, respectively. Applying some pruning strategy, only a subset of the 1-feature-value pairs will be kept and each 1-feature-value pair forms an associative classification rule denoted as $A_m^i \Rightarrow C_j$. Unlike the traditional AC that the itemsets and rules are pruned based on the frequency count (support) and accuracy (confidence) respectively, our MCA-based associative rules are generated based on some measure of the correspondence (to be discussed next).

2.3. Multiple Correspondence Analysis (MCA)

Multimedia data is projected into a new space using the first and second principle components [10]. The pseudo-code for generating the 1-feature-value pair rules is presented as follows, where J is the total number of classes, K is the total number of 1-feature-value pairs, and R is the final rule set.

1-FEATURE-VALUE PAIR RULE GENERATION

$R \leftarrow \emptyset$;
for $j \leftarrow 1$ **to** J
  for $k \leftarrow 1$ **to** K
    $angle_k = cos^{-1} \frac{<pair_k, class_j>}{||pair_k|| * ||class_j||}$;
    if $angle_k < threshold1$
    then $R \leftarrow R \cup \{pair_k \Rightarrow class_j\}$.

In our proposed framework, the inner product of the feature-value pairs and classes are used to represent the similarity, and the angle between the feature-value pairs and classes is used as a measurement [10]. In addition, the feature-value pairs whose corresponding angle values were smaller than the angle threshold were selected for classification rules. In this paper, a different method is developed to determine the threshold for angles. Rather than computing the average angle values in certain ranges, we evaluate the rules by applying different thresholds to the training set and determine the threshold1 value which yields the highest accuracy.

After 1-feature-value pairs are generated, 2-feature-value pairs are generated by pairing the 1-feature-value pairs. However, unlike the traditional AC, the pairing depends on whether both of the 1-feature-value pair rules have been generated for the same class. The pseudo-code for generating a 2-feature-value pair rule is as follows. Here, J is the total number of classes and S_j is the total number of selected 1-feature-value pairs for class C_j. The value of threshold2 is determined similar to threshold1 based on the training set.

2-FEATURE-VALUE PAIR RULE GENERATION

for $j \leftarrow 1$ **to** J
  for $s \leftarrow 1$ **to** $S_j - 1$
    for $t \leftarrow s + 1$ **to** S_j
      if $pair_s$ and $pair_t$ do not belong to the same feature **then**
        $angle_{st} = cos^{-1} \frac{<\{pair_s, pair_t\}, class_j>}{||\{pair_s, pair_t\}|| * ||class_j||}$;
        if $angle_{st} < threshold2$ **then**
          $R \leftarrow R \cup \{pair_s \wedge pair_t \Rightarrow class_j\}$.

3. EXPERIMENTS AND RESULTS

Our proposed framework is evaluated with the sampled TRECVID 2007 and 2008 videos. We use all the positive instances in our database, and randomly choose the negative instances. The concepts used include hand, urban, crowd, person, two-people, outdoor, building, vegetation, and road, whose descriptions could be found in [12]. The reason we selected these concepts only because there are sufficient amounts of instances to build useful training and testing sets in our database. The precision, recall, and F1-score metrics are used under the 3-fold cross validation approach as explained earlier. To show the efficiency of our proposed framework, its performance is compared to the Decision Tree classifier (DT), Support Vector Machine classifier (SVM), and Association Rule Classification (ARC) available in WEKA [1]. Note that we use the default values for parameters of three classifiers WEKA has provided. The average precision (Pre), recall (Rec), and F1-score (F1) values obtained over the three folds are presented in Tables 1 and 2, where columns 2 to 4 provide the performance of WEKA's DT, SVM, and ARC, respectively, and the last column provides the performance of our proposed framework.

Hand	**DT**	**SVM**	**ARC**	**MCA**
Pre	0.33	0.46	0.28	0.40
Rec	0.06	0.31	0.20	0.80
F1	0.10	0.37	0.24	0.53
Urban	**DT**	**SVM**	**ARC**	**MCA**
Pre	0.53	0.51	0.49	0.45
Rec	0.25	0.41	0.41	0.76
F1	0.34	0.46	0.43	0.57
Crowd	**DT**	**SVM**	**ARC**	**MCA**
Pre	0.78	0.54	0.40	0.41
Rec	0.03	0.32	0.41	0.83
F1	0.06	0.40	0.39	0.55
Person	**DT**	**SVM**	**ARC**	**MCA**
Pre	0.61	0.54	0.49	0.45
Rec	0.32	0.39	0.32	0.71
F1	0.42	0.45	0.39	0.55

Table 1. Performance evaluation for four concepts

Two people	DT	SVM	ARC	MCA
Pre	0.00	0.51	0.14	0.37
Rec	0.00	0.19	0.09	0.86
F1	0.00	0.27	0.11	0.51
Outdoor	**DT**	**SVM**	**ARC**	**MCA**
Pre	0.59	0.59	0.42	0.47
Rec	0.37	0.39	0.60	0.73
F1	0.46	0.47	0.47	0.57
Building	**DT**	**SVM**	**ARC**	**MCA**
Pre	0.55	0.57	0.45	0.43
Rec	0.27	0.34	0.34	0.83
F1	0.36	0.42	0.38	0.57
Vegetation	**DT**	**SVM**	**ARC**	**MCA**
Pre	0.54	0.51	0.48	0.41
Rec	0.14	0.35	0.35	0.81
F1	0.21	0.42	0.41	0.54
Road	**DT**	**SVM**	**ARC**	**MCA**
Pre	0.59	0.58	0.49	0.43
Rec	0.35	0.34	0.48	0.87
F1	0.44	0.42	0.47	0.58

Table 2. Performance evaluation for five concepts

As can be seen from both tables, our proposed concept detection framework achieves promising results compared to DT, SVM, and ARC. In fact, our recall values outperform the other three classifiers, and our F1-scores can achieve at least 10% higher than the ones obtained by the other three classifiers. This demonstrates that the proposed novel associative classification algorithm with MCA is effective in video concept detection.

4. CONCLUSIONS

In this paper, a novel associative classification algorithm using MCA for video concept detection is proposed. MCA is utilized to measure the correlation between different feature-value pairs and classes to infer the high-level concepts from the extracted low-level features. Nine concepts from the TRECVID 2007 and 2008 data are used to validate our proposed framework. The experimental results demonstrate that our proposed framework achieves promising concept detection results. Furthermore, we are able to show an increased overall performance over some well-known classifiers.

5. REFERENCES

[1] I. H. Witten and E. Frank, *Data Mining: Practical Machine Learning Tools and Techniques*, Morgan Kaufmann, second edition, June 2005.

[2] I. Bouzouitz and S. Elloumi, "Integrated generic association rule based classifier," in *IEEE International Conference on Database and Expert Systems Applications (DEXA07)*, September 2007, pp. 514–518.

[3] F. Thabtah, "Challenges and interesting research directions in associative classification," in *IEEE International Conference on Data Mining Workshops (ICDMW06)*, December 2006, pp. 785–792.

[4] 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.

[5] 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.

[6] B. Liu, W. Hsu, and Y. Ma, "Integraing classification and association rule mining," in *International Conference on Knowledge Discovery and Data Mining (KDD98)*, August 1998, pp. 80–86.

[7] C. S. Kanimozhih Selvi and A. Tamilarasi, "Association rule mining with dynamic adaptive support thresholds for associative classification," in *IEEE International Conference on Computational Intelligence and Multimedia Applications (ICCIMA07)*, December 2007, pp. 76–80.

[8] P. Vateekul and M.-L. Shyu, "A conflict-based confidence measure for associative classification," in *IEEE International Conference on Information Reuse and Integration (IRI08)*, July 2008, pp. 256–261.

[9] X. Yin and J. Han, "CPAR: Classification based on predictive association rules," in *SIAM International Conference on Data Mining (SDM03)*, May 2003, pp. 369–376.

[10] L. Lin, G. Ravitz, M.-L. Shyu, and S.-C. Chen, "Correlation-based video semantic concept detection using multiple correspondence analysis," in *IEEE International Symposium on Multimedia (ISM08)*, December 2008, pp. 316–321.

[11] N. J. Salkind, Ed., *Encyclopedia of Measurement and Statistics*, Thousand Oaks, CA: Sage Publications, Inc, 2007.

[12] 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.