Association rule mining with a correlation-based interestingness measure for video semantic concept detection

Lin Lin, Mei-Ling Shyu*

Department of Electrical and Computer Engineering
University of Miami, Coral Gables, FL 33124, USA
E-mail: l.lin2@umiami.edu, shyu@miami.edu
*Corresponding author

Shu-Ching Chen

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

Abstract: Content-based multimedia retrieval and automatic semantic concept detection research areas have been motivated by the high demands of multimedia applications and services. Due to its high efficiency and good performance, association rule mining (ARM) has been adopted to discover the association patterns from the multimedia data and predict the target concept classes in various media types. As a rule-based method, ARM faces the challenges on rule pruning in both rule generation and rule selection stages. Such challenges could be addressed by utilizing proper interestingness measures, and therefore an interestingness measure plays an important role in association rule mining and multimedia retrieval research. In this paper, a video semantic concept detection framework that uses ARM together with a novel correlation-based interestingness measure is proposed. The interestingness measure is obtained from applying multiple correspondence analysis (MCA) to capture the correlation between the features and concept classes and to bridge the semantic gap between low-level features and high-level concepts. This new correlation-based interestingness measure is first used in the rule generation stage, and then reused and combined with the inter-similarity and intra-similarity values to select the final rule set for classification. Experimented with 14 high-level concepts from the benchmark data provided by the TRECVID project, our proposed framework achieves higher accuracy than the other six classifiers that are commonly used in multimedia retrieval and concept detection.

Keywords: Association rule mining (ARM); Multiple correspondence analysis (MCA); Interestingness measure; Semantic concept detection.

Reference to this paper should be made as follows: Lin, L., Shyu, M.-L. and Chen, S.-C. (2010) 'Association rule mining with a correlation-based interestingness measure for video semantic concept detection', *Int. J. Information and Decision Sciences*, Vol. x, No. x, pp.xxx–xxx.

Biographical notes: Lin Lin received her Master degree in Electrical and Systems Engineering from University of Pennsylvania, Philadelphia, Pennsylvania, USA in 2006. She is currently pursuing her Ph.D. degree at the Department of Electrical and Computing Engineering from the University of Miami, Florida, USA. Her research interests include multimedia retrieval, multimedia databases, and multimedia data mining. She received the "Best Student Paper Award" from the Third IEEE International Conference on Semantic Computing (ICSC) in September 2009.

Dr. Mei-Ling Shyu has been an Associate Professor at the Department of Electrical and Computer Engineering (ECE), University of Miami (UM) since June 2005. Prior to that, she was an Assistant Professor in ECE at UM dating from January 2000. She received her Ph.D. degree from the School of Electrical and Computer Engineering, Purdue University, West Lafayette, IN, USA in 1999, and her three master degrees from Computer Science, Electrical Engineering, and Restaurant, Hotel, Institutional, and Tourism Management from Purdue University in 1992, 1995, and 1997. Her research interests include multimedia data mining and multimedia information systems.

Dr. Shu-Ching Chen is a Full Professor in the School of Computing and Information Sciences (SCIS), Florida International University (FIU), Miami since August 2009. Prior to that, he was an Assistant/Associate Professor in SCIS at FIU from 1999. He received Master's degrees in Computer Science, Electrical Engineering, and Civil Engineering in 1992, 1995, and 1996, respectively, and the Ph.D. degree in Electrical and Computer Engineering in 1998, all from Purdue University, West Lafayette, IN, USA. His main research interests include distributed multimedia database management systems and multimedia data mining.

1 Introduction

Motivated by a large number of requirements in multimedia retrieval applications and services, including the sports video summarizers, content-based video search engines, personalized video collections, etc., automatic video semantic concept detection has been rapidly developed (Lew et al., 2006; Snoek and Worring, 2008). For example, detecting the high-level concepts such as face and hand from the video camera would help in security check, building and street from video games could be used for navigation, and the scenes with urban and waterscape would assist the users who are planning a trip. Moreover, some users may be interested in the collection of the videos with the vegetation or sky; some users may search for the videos with the semantics of a person or the crowd. Classification using association rule mining (ARM) for video semantic concept detection has been studied (Lin et al., 2007; Zhao et al., 2007), taking advantages of both classification and ARM techniques.

There are two main stages in classification using ARM (Liu et al., 1998). The first one is the stage of association rule discovery or rule generation, and the other one is the stage of association rule selection or rule ranking. As a rule-based method, the ARM approaches face the challenges on rule pruning (Thabtah, 2006) in both stages in order to reduce the time and space cost in the mining process. To address these challenges, effective interestingness measures, one of the core elements in generating rules and selecting rules (properly pruning the rules), can be utilized. In the rule generation stage, the interestingness measures are used to discover the feature-value pairs and remove the un-interesting feature-value pairs. The mining efficiency could be improved in this stage by using the frequency count (support) in the traditional ARM algorithms. In the rule selection stage, the interestingness measures are used to rank the candidate rules and prune the low-ranked rules. The mining accuracy could be improved in this stage by applying the evaluation criterion (confidence) in the traditional ARM algorithms. One of the well-known algorithms is Apriori (Agrawal and Srikant, 1994) which utilizes the interestingness measures to discover the frequent feature-value pairs satisfying the minimum support value and select strong rules satisfying the minimum confidence value.

In this paper, a novel video semantic concept detection framework facilitated with a correlation-based interestingness measure for association rule generation and selection is proposed. MCA (multiple correspondence analysis) is utilized to evaluate each of the extracted low-level features and the classes and then generate the 2-feature-value pairs as the rules that better represent each investigated concept. Next, the correlation information obtained from MCA in the previous stage is reused and aggregated with the inter-similarity and intra-similarity values of the rules to rank the candidate rules. The selected rule are used for classification that the concept class is determined by the majority class of the matched rules. To evaluate our proposed framework, the high-level concepts and videos from TRECVID 2007 and 2008 (Smeaton et al., 2006) are used, and the performance is compared with the well-known *decision tree* classifier, *support vector machine* classifier, *Neural Network* classifier, *Kth Nearest Neighbor* classifier, *AdaBoost* classifier, and *one rule based JRip* classifier. Overall, our proposed framework outperforms all six classifiers on both recall and $F1$-score values which are commonly used criteria for accuracy evaluation.

The paper is organized as follows. Section 2 presents the related work utilizing different interestingness measures for ARM. In Section 3, the proposed framework is presented and detailed discussions on its different components are provided. Section 4 discusses the experiments as well as the analyses of the results. The paper is then concluded in Section 5 and future work is also discussed in this section.

2 Related Work

Typically, there are three categorized interestingness measures for generating and selecting rules: objective measures, subjective measures, and semantic measures (Geng and Hamilton, 2006). The objective measures are calculated based on probability, statistics, distance, or information theory. Most of the criteria depend only on the data characteristics, such as conciseness, coverage, reliability, peculiarity, and diversity. In (Malik and Kender, 2006), the objective measure is applied to prune the rules for web image clustering. First, the visual features from images and textual features from web pages were extracted. Next, association rules were generated by using an improved Apriori algorithm, and various statistically-inspired interestingness measures were evaluated by their abilities of pruning the generated rules. Last, the hypergraphs were generated from the rules, and a hypergraph partitioning algorithm was used to assign images to different clusters. In most of the studies, the objective interestingness measures were used to select one or several proper measures based on their properties or by an interactive manner. The authors in (Nguyen et al., 2008) introduced an approach to aggregate a set of objective interestingness measures using the Choquet integral as the aggregation operator to find the most interesting association rules.

The subjective measures such as surprisingness, unexpectedness, and novelty consider both the data and the user's domain knowledge about the data. In (Liu et al., 2000), an interestingness analysis system (IAS) was developed to assist the users in finding unexpected rules from a set of discovered association rules. Unexpectedness means that those rules are interesting if they are unknown to the user or contradict the user's expectations. The proposed IAS leveraged the user's existing domain knowledge to analyze discovered associations and then ranked the discovered rules according to various interestingness criteria, such as conformity and various types of unexpectedness. Yu et al. (2003) introduced a ShotRank notion as a measure of subjective interestingness for a video browsing and summarization system. The system utilized previous viewers' browsing log to facilitate future viewers and applied an interestingness measure to unify video analysis and user browsing log mining. Experimental results showed that ShotRank was able to represent the subjective notion of interestingness of each video shot and improve the future viewers' browsing experience.

The semantic measures take into account the semantics and explanations of the feature-value pairs, such as utility and actionability considering the semantics of the data. The utility-based measures consider not only the statistical aspects of the raw data but also the utility of the mined patterns. Actionability-based measures take into consideration that the rules are interesting if the users can do something with them to their advantage. In (Lin et al., 2008), we have introduced the utilization of *Multiple Correspondence Analysis (MCA)* as a utility-based semantic measure for association rule generation. MCA is an extended approach of correspondence analysis (CA). Traditional CA is a descriptive data analytic technique designed to analyze simple two-way tables, containing some measure of correspondence between the rows and columns. By using MCA, multi-way tables for more than two variables could be analyzed, so that the correspondence between the features and classes (columns) through the instances (rows) could be explored in a multimedia database. In this paper, the 2-feature-value pair association rules are generated by a similar approach that was applied to generate the 1-feature-value pair association rules in (Lin et al., 2008), but the focus is on selecting those 2-feature-value pair association rules.

Selecting the association rules is one of the challenging aspects in classification using ARM and thus becomes a popular research topic. In order to select more interesting rules, research work on ranking strategies has been conducted. There are various approaches that rank the association rules using the interestingness measures. Most of the algorithms are investigated based on traditional support and confidence values or based on improved support and confidence values. The authors in (Song et al., 2006) proposed a rule ranking strategy where the length of the rules has the highest priority, followed by the confidence value, and then the support value. That is, association rules with the same length were sorted according to their confidence values, and if the rules have the same length and the same confidence value, they were ranked based on their support values. More constraints to the rule ranking methods rather than using the confidence and support values only were proposed in (Rodda and Shashi, 2007). If the rules have the same confidence and support values, more general rules were preferred over the specific rules. Then the rules with better simplicity were selected if all the above constraints were the same, and the last constrain was based on which rule was generated earlier. A personalized association rule ranking method based on semantic similarity was proposed in (Yang et al., 2008). In addition to the statistical information such as support, confidence, and chi-square values, the rules were also ranked and selected by the similarity between the rules and the keywords assigned by the user's interests.

Rather than using the tradition support and confidence definitions, some new support and confidence methods were defined. A confidence gain measure for association rule scoring was introduced in (Tamir and Singer, 2006), which combines both the confidence measure and lift measure, taking advantages of both measures. The confidence gain presented the local confidence of a feature-value pair compared to its average confidence in the given database, and outperformed several association measures in the experiments. The authors in (Zhu et al., 2008) redefined the support and confidence values and mined the cross-modal association rules that associated keywords with several visual feature clusters. The images in the interesting clusters were ranked higher, and the clusters in these rules can be sorted by their corresponding confidence values in the descending order. In (Vateekul and Shyu, 2008), a novel conflict-based confidence measure with an interleaving ranking strategy for re-ranking association rules was proposed. The new confidence measure captured the inter-distance between a rule and a training data instance, and the experiments showed that their proposed framework achieved better performance than the traditional confidence measures for both balanced and imbalanced data sets.

Furthermore, some other strategies were proposed for rule ranking and selection. In (Fogarty et al., 2008), CueFlik which is a Web image search application allowed the users to create their own rules to re-rank any future Web image search results according to visual characteristics. The rules were ranked by rule scores which were the distance between the future image and each positive or negative example dividing the distance to the nearest positive example by the sum of the distances to the nearest positive and nearest negative examples. In (Liu et al., 2008), the relationship between high-level concepts was discovered by using ARM. The co-occurrence of several semantic concepts could imply the presence of other concepts. The prediction value of the detector indicated the likelihood that the detector regarded the presence of a certain concept. The association rules were ranked by using the combination of associations with the prediction values. In the experiments on TRECVID 2005 data, the authors showed that the detection accuracy was improved.

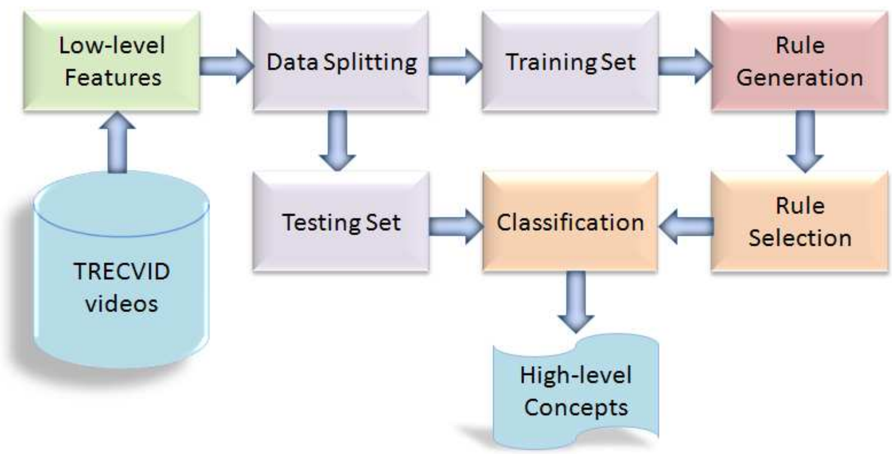

Figure 1 The Proposed Framework

3 Association Rule Mining with a Correlation-based Interestingness Measure

This paper proposes a novel framework that performs video semantic concept detection (classification) via the use of a correlation-based interestingness measure for association rule generation and the reuse of the correlation-based interestingness measure together with the inter-similarity and intra-similarity values for rule selection. Our proposed framework is shown in Figure 1, which is composed of the following four stages.

1. Low-level feature extraction (to be discussed in Section 3.1);
2. Splitting data to training set and testing set (to be discussed in Section 3.1);
3. Rule generation (to be discussed in Section 3.2); and
4. Rule selection and classification (to be discussed in Section 3.3).

3.1 Low-level Feature Extraction and Data Splitting

The details of stage 1 and stage 2 are shown in Figure 2. Note that since shot boundary information has been provided by TRECVID (Smeaton et al., 2006), segmentation is beyond the scope of this paper and the extracted audio-visual features are shot-based. At the first stage, 16 audio features, 11 visual features, and 1 meta feature (i.e., the length of the shot) are extracted. The normalization process is applied to scale all continuous values (except the class label) per video. The normalization method is to subtract the minimum value and divide by the distance between the maximum and the minimum values for each feature, so that the values of each feature in the data set lie between zero and one. Then the data instances in a multimedia database are characterized by $F+1$ low-level attributes/features/columns, i.e., F normalized numerical features A_f (where f=1 to F) and 1 nominal class label C_j (where C_j=C_p or C_n, C_p is the target concept class, and C_n is the non-target concept class) as shown in Table 1.

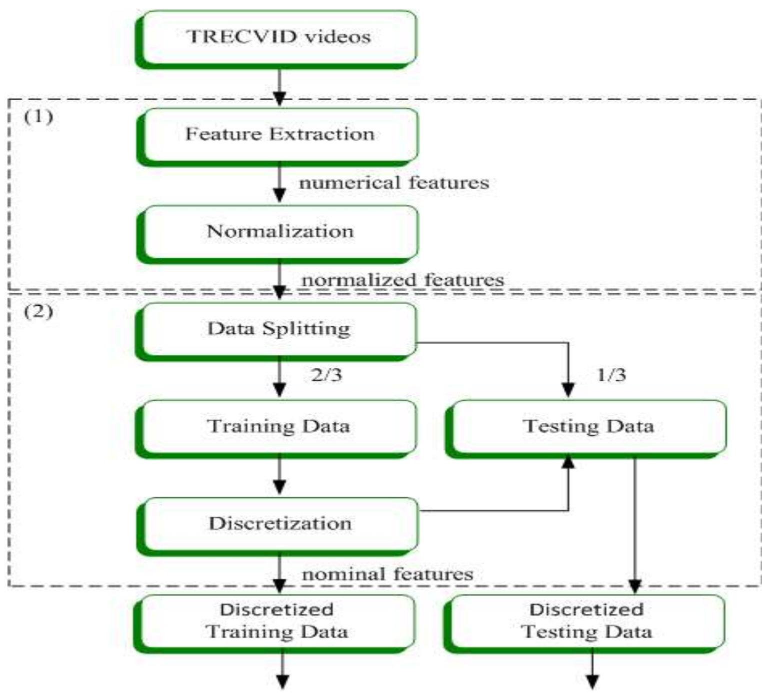

Figure 2 The Details of Stage 1 and Stage 2

The second stage is to split the data instances into a training set and a testing set. For each concept, the data instances are first split to two sets, namely (i) the positive set including all the data instances labeled with the target concept class (i.e., positive instances) and (ii) the negative set including all the data instances labeled with the non-concept class (i.e., negative instances). Two-third of the positive set and negative set are randomly selected and combined as the training data, and the remainder one-third of the positive set and negative set are used as the testing data. The main reasons of this process are that (1) the splitting is able to mitigate any bias caused by the particularly chosen instances, (2) each class being properly represented in both training and testing sets could be guaranteed, and (3) the 3-fold stratified cross-validation approach ensures that each data instance could be tested.

Due to the fact that ARM requires the input data to be nominal, all the extracted features are discretized at this stage before training the model. The methods used for discretization include (i) the information gain method introduced in (Fayyad and Irani, 1992), and (ii) the disparity measure using the average value

Table 1 Example data instances in the multimedia database after normalization

$feature_1$	$feature_2$	...	$feature_F$	$class_j$
0.23	0.38	...	0.15	C_p
0.17	0.67	...	0.02	C_n
0.15	0.78	...	0.16	C_p
...	...	...	...	...
0.10	0.59	...	0.84	C_n
...	...	...	...	...

Table 2 Example data instances in the multimedia database after discretization

$feature_1$	$feature_2$	...	$feature_F$	$class_j$
A_1^2	A_2^2	...	A_F^2	C_p
A_1^1	A_2^3	...	A_F^1	C_n
A_1^1	A_2^3	...	A_F^2	C_p
...	...	...	...	...
A_1^1	A_2^3	...	A_F^4	C_n
...	...	...	...	...

of the feature to construct two partitions for that feature. The second method is applied when the first method fails to generate more than one partition. The training data set is discretized into various partitions and the same ranges of these partitions are applied to discretize the testing data set. These partitions generated by the discretization process are called *feature-value pairs* in our study. By applying discretization, each feature A_f has several possible nominal feature-value pairs A_f^i (where i=1 to K_f and $\sum K_f = K$). For instance, A_{17} is the feature of pixel changes, which is converted to 3 partitions (i.e., $K_{17} = 3$), and A_{17}^1, A_{17}^2, and A_{17}^3 represent the partitions of the feature value ranges [0, 0.32865], (0.32865, 0.5044], and (0.5044, 1], respectively. Table 2 presents some example discretized data instances.

3.2 Rule Generation

The details of stage 3 are shown in Figure 3. The combination of each two 1-feature-value pairs that do not belong to the same feature is considered as a 2-feature-value pair. 2-feature-value pairs are represented by $\{A_{f1}^{i1}, A_{f2}^{i2}\}$, where $f1$, $f2$=1 to F, $f1 \neq f2$, $i1$=1 to K_{f1}, $i2$=1 to K_{f2}, and the classes are C_p as the target concept class and C_n as the non-concept class.

From the traditional ARM algorithm, both rules $A_1^1 \wedge A_2^3 \Rightarrow C_p$ and $A_1^1 \wedge A_2^3 \Rightarrow C_n$ might be generated (as shown in Table 2). This indicates that for example, by using the frequency count, the 2-feature-value pairs $\{A_1^1, A_2^3\}$ might represent both C_p and C_n. However, this is conflicting in classification. In (Lin et al., 2008), the utilization of MCA to analyze the multimedia data instances described by a set of low-level features and high-level concepts was explored. The study showed that for each 1-feature-value pair, it will be classified to only one class which has the larger correlation. Therefore, all the 1-feature-value pairs might represent either C_p or C_n, but not both (i.e., exclusively). The other advantage taking from

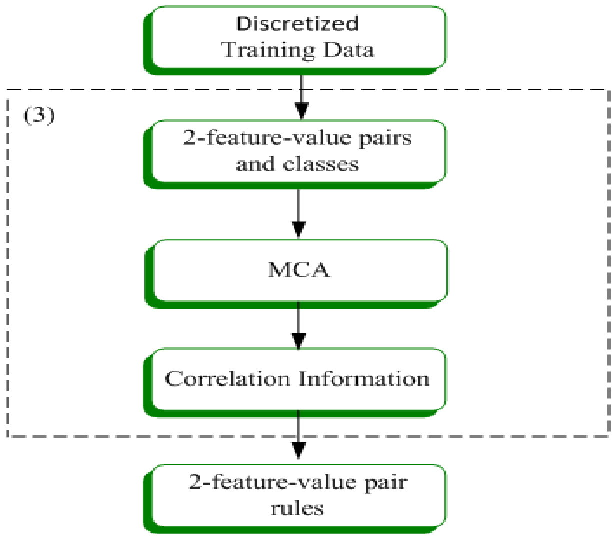

Figure 3 The Details of Stage 3: Rule Generation

MCA is that MCA is the extension of traditional correspondence analysis so that it has the ability to analyze tables containing some measure of correspondence between the rows and columns with multiple variables (Salkind, 2007). Therefore, the correlation information between the 2-feature-value pairs ($\{A^{i1}_{f1}, A^{i2}_{f2}\}$) and classes ($C_p$ and C_n) can be calculated by applying MCA to the discretized training data set.

Assume that there are N data instances in the multimedia database and the total number of 2-feature-value pairs is S. MCA codes the data by creating a binary column with the constraint that one and only one of the columns gets the value 1 for each nominal variable (i.e., 2-feature-value pair). This results in a matrix which is called the indicator matrix X with size $N \times S$. Rather than analyzing the indicator matrix as in traditional CA, the inner product of the indicator matrix called Burt matrix Y (of size $S \times S$) is analyzed in MCA. Now, let the grand total of the Burt matrix be G. The probability matrix Z, the mass matrix M (of size $1 \times S$), and the main diagonal of the mass matrix D could be captured. MCA will provide the principle components from singular value decomposition (SVD) as shown in Equation (1).

$$D^{-\frac{1}{2}}(Z - MM^T)(D^T)^{-\frac{1}{2}} = P\Delta Q^T, where \tag{1}$$

- $Z = Y/G$ and $Y = X^T X$;
- M is the vector of the column totals of Z and $D = diag(M)$;
- Δ is the diagonal matrix of the singular values;
- Columns of P are the left singular vectors (gene coefficient vectors) in SVD;
- Rows of Q^T are the right singular vectors (expression level vectors) in SVD.

The feature-value pairs and the classes can then be projected into a new space using the first and second principle components. The inner product of all possible 2-feature-value pairs and the classes are calculated, and the angles between the 2-feature-value pairs and the classes are used as a measurement to represent the correlation. It is clear that the higher correlation between a feature-value pair and a class, the smaller angle value between them.

The pseudo-code for computing the angle values and generating a 2-feature-value pair rule is presented as follows. In our proposed framework, we keep those 2-feature-value pairs whose angle values are smaller than an angle threshold. To determine the angle threshold automatically, different thresholds are applied to the feature-value pairs (i.e., the range between 5 degrees and 90 degrees with 5 degrees increase in each step) and the classification accuracy of the training set is evaluated, among which the one having the highest accuracy is considered as the threshold value. The same threshold is used for the positive and negative feature-value pair sets so that the computational cost can be reduced.

2-Feautre-Value Pair Rule Generation

```
k ← 0;
l ← 0;
for class_r ← class_p to class_n
    for s ← 1 to S
        angle_s = cos^{-1} <pair_s,class_r> / (||pair_s||*||class_r||);
        if angle_s < threshold then
            rule_i ← pair_s ⇒ class_r;
            if class_r = class_p then
                k ← k + 1.
            else if class_r = class_n then
                l ← l + 1.
    end
end
```

Let $f1$ and $f2$=1 to F, $f1 \neq f2$, $i1$=1 to K_{f1}, and $i2$=1 to K_{f2}. Equation (2) denotes the rules for class C (C_p for the target concept class or C_n for the non-concept class).

$$A_{f1}^{i1} \wedge A_{f2}^{i2} \Rightarrow C. \tag{2}$$

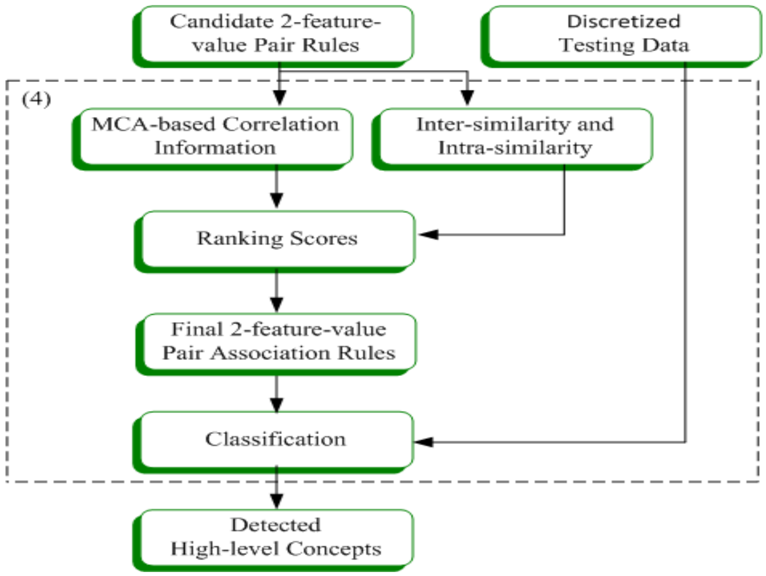

Figure 4 The Details of Stage 4: Rule Selection and Classification

3.3 Rule Selection and Classification

After generating the 2-feature-value pair rules (as in the format of Equation (2)), the correlation-based interestingness measure that reuses the correlation information is converted to the score value. Moreover, the inter-similarity and intra-similarity values are calculated and integrated as the similarity information to evaluate each rule. The inter-similarity and intra-similarity are defined as follows. For each candidate rule, the inter-similarity is the similarity between the rule and the data instances which have different class labels, and the intra-similarity is the similarity between the rule and the data instances which have the same class labels. The details of stage 4 are shown in Figure 4.

Let the number of positive instances be N_{pos}, the number of negative instances be N_{neg}, the number of candidate 2-feature-value pair association rules generated for the target concept be R_{pos}, and the number of the candidate 2-feature-value pair association rules for non-concept class be R_{neg}. We define the intra-similarity and inter-similarity measures of a rule as follows.

$$IntraSimilarity\_k = \quad \alpha_1 \sum_{counter1} 1 + \alpha_2 \sum_{counter2} 1 + \alpha_3 \sum_{counter3} 1 \quad /N_{pos}, \tag{3}$$

$$IntraSimilarity\_l = \quad \alpha_1 \sum_{counter1} 1 + \alpha_2 \sum_{counter2} 1 + \alpha_3 \sum_{counter3} 1 \quad /N_{neg}, \tag{4}$$

where

- $k = 1$ to R_{pos} and $l = 1$ to R_{neg};
- $\alpha_1 = 1$, $\alpha_3 = 0$, $\alpha_2 = r/2$, and r is the ratio of the number of positive instances to the number of negative instances;
- $counter1$ is the counter of the event that both two feature-value pairs are matched;

- $counter2$ is the counter when only one feature-value pair is matched;
- $counter3$ is the counter that none of the feature-value pairs is matched.

$$InterSimilarity\_k = \beta_1 \sum_{counter1} 1 + \beta_2 \sum_{counter2} 1 + \beta_3 \sum_{counter3} 1 \;\; /N_{pos}. \quad (5)$$

$$InterSimilarity\_l = \beta_1 \sum_{counter1} 1 + \beta_2 \sum_{counter2} 1 + \beta_3 \sum_{counter3} 1 \;\; /N_{neg}. \quad (6)$$

where

- $k = 1$ to R_{pos} and $l = 1$ to R_{neg},
- $\beta_1 = 0,\ \beta_3 = 1,\ \beta_2 = 1 - r/2$ and r is the ratio of the number of positive instances to the number of negative instances
- $counter1$ is the counter of the event that both two feature-value pairs are matched,
- $counter2$ is the counter when only one feature-value pair is matched,
- $counter3$ is the counter that none of the feature-value pair is matched.

Please note that based on our definitions, a larger intra-similarity value ($\in [0, 1]$) indicates a better rule and a larger inter-similarity value ($\in [0, 1]$) also indicates a better rule. Therefore, we can calculate the sum of these inter-similarity and intra-similarity values ($\in [0, 2]$), which serves as a similarity-based score value. In this manner, the inter-similarity and intra-similarity values contribute equally to the similarity information.

In this stage, the reused correlation information is converted to the correlation-based score values from the angle values captured from the rule generation stage using Equation (7). Please note that from this definition, larger A_k and A_l values ($\in [0, 1]$) also indicate better rules.

$$\begin{aligned} A_k &= (1 - angle_k/90); \\ A_l &= (1 - angle_l/90). \end{aligned} \quad (7)$$

where

- $k = 1$ to R_{pos} and $l = 1$ to R_{neg};
- $angle_k$ is the angle value of the k^{th} candidate positive feature-value pair;
- $angle_l$ is the angle value of the l^{th} candidate negative feature-value pair.

The correlation information is aggregated with the similarity information to calculate the final rule ranking scores. That is, the final ranking score is the sum of the correlation-based score and the similarity-based score. Similarly, in this manner, the correlation information and the similarity information have an equal contribution to the ranking strategy. A threshold is set so that the 2-feature-value pair rules whose final ranking scores are larger than the threshold value are selected as the finial rule set; otherwise the rules are removed. The threshold is selected as the one which yields the highest accuracy when applying the generated rule set with various thresholds in the training process.

Now, the final association rule set consists of all the selected 2-feature-value pair rules and is used for classification for each concept (class label). The steps are: (1) Each data instance in the testing set is checked to see if it consists of any of the 2-feature-value pairs from the selected rules; (2) If the 2-feature-value pairs exist in the testing data instance, the class labels of the corresponding matched 2-feature-value pairs are collected and the majority class is assigned to the testing data instance; and (3) If there are an equal number of matched positive rules and negative rules, then the label of the data instance is set to positive. For each concept, these three steps are repeated. In other words, different 2-feature-value pair rules will be selected as the final association rule sets for each concept.

4 Performance Evaluation

To validate our proposed framework, the videos available for the TRECVID 2007 and 2008 high-level feature extraction task are used. In the experiments, 14 concepts such as two-people ($c7$), outdoor ($c8$), building ($c9$), vegetation ($c11$), street ($c12$), road ($c13$), sky ($c14$), hand ($c15$), urban ($c16$), waterscape ($c17$), crowd ($c18$), face ($c19$), person ($c20$), and walking ($c34$) are used for performance evaluation. The descriptions of these concepts can be found in (Smeaton et al., 2006).

The performance of our proposed framework is compared to those of the *decision tree* classifier using $C4.5$ algorithm (DT), *support vector machine* classifier using Sequential Minimal Optimization (SMO) algorithm (SVM), *Neural Network* classifier using Multilayer Perceptron algorithm (MP), *Kth Nearest Neighbor* classifier using IbK algorithm (KNN), *AdaBoost* (ADA) classifier, and *one rule based JRip* classifier (JR). These classifiers are available in WEKA (Witten and Frank, 2005). In the experiments, the default parameters in WEKA are adopted and the performance metrics (including the average precision, recall, and F1-score values obtained over the three folds) are used. The results are presented in Table 3 and Table 4, where columns 2 to 7 provide the performance of WEKA's DT, SVM, MP, KNN, ADA, and JR respectively, and the last column provides the performance of our proposed framework.

As can be seen from Table 3 and Table 4, our proposed video semantic concept detection framework using ARM together with a new correlation-based interestingness measure outperforms the DT, SVM, MP, KNN, ADA, and JR classifiers, in both the recall values and $F1$-scores. Moreover, it can be observed from the result tables that some of the compared classifiers perform well for certain concept classes, but not for all the concept classes. For instance, the JR classifier gives the second best $F1$-scores on the **street**, **urban**, and **walking** concepts, the ADA classifier gives the second best $F1$-score on the **road** concept, and the MP classifier gives the second best $F1$-score on the **crowd** concept. However, none of them achieves the same performance for all the 14 investigated concepts. On the other hand, our proposed framework demonstrates that it performs the best in comparison to all the other 6 classifiers in all 14 investigated high-level concepts. Note that for SVM classifier, it fails for some extremely imbalanced concepts as discussed in (Lin et al., 2008).

Table 3 Performance evaluation for seven concepts

concept	evaluation	**DT**	**SVM**	**MP**	**KNN**	**ADA**	**JR**	**ARM**
Two-people	Precision	0.51	0.00	0.44	0.40	0.40	0.52	0.36
	Recall	0.19	0.00	0.25	0.47	0.36	0.19	0.92
	F1-score	0.27	0.00	0.32	0.43	0.38	0.11	0.51
Outdoor	Precision	0.51	0.00	0.44	0.51	0.52	0.58	0.46
	Recall	0.19	0.00	0.25	0.53	0.51	0.46	0.75
	F1-score	0.27	0.00	0.32	0.52	0.51	0.51	0.56
Building	Precision	0.51	0.00	0.44	0.47	0.51	0.52	0.44
	Recall	0.19	0.00	0.25	0.54	0.46	0.40	0.80
	F1-score	0.27	0.00	0.32	0.50	0.48	0.45	0.57
Vegetation	Precision	0.57	0.55	0.50	0.44	0.46	0.53	0.43
	Recall	0.34	0.27	0.42	0.49	0.42	0.34	0.67
	F1-score	0.42	0.36	0.46	0.46	0.44	0.41	0.52
Street	Precision	0.55	0.58	0.49	0.47	0.55	0.54	0.51
	Recall	0.50	0.47	0.49	0.58	0.58	0.53	0.74
	F1-score	0.52	0.52	0.49	0.52	0.52	0.54	0.60
Road	Precision	0.58	0.59	0.50	0.49	0.54	0.57	0.46
	Recall	0.34	0.35	0.47	0.58	0.49	0.44	0.81
	F1-score	0.42	0.44	0.48	0.53	0.51	0.49	0.59
Sky	Precision	0.62	0.64	0.56	0.53	0.55	0.67	0.47
	Recall	0.48	0.45	0.51	0.56	0.52	0.45	0.79
	F1-score	0.54	0.52	0.54	0.54	0.54	0.52	0.59

For further performance evaluation, Figure 5, Figure 6, and Table 5 are generated. It can be easily seen that our proposed framework achieves at least 25% to at most 48% average improvement for the recall values over all concepts, and at least 7% to at most 21% average improvement for the $F1$-score values are gained over all concepts. The better recall values (overall 35% improvement over all classifiers) demonstrate that more target concept instances are classified correctly. The better $F1$-scores mean that the recall values (overall 11% improvement over all classifiers) increase without compromising the precision values too much. This observation clearly shows the superior performance of our proposed framework with the new interestingness measure in both rule generation and rule selection stages.

5 Conclusion and Future Work

Semantic concept detection has become a popular research area, motivated by the high demands on multimedia applications and services dealing with large amounts of multimedia data. The association rule mining (ARM) technique has also commonly utilized in multimedia retrieval and concept detection due to its

Table 4 Performance evaluation for seven concepts

concept	evaluation	**DT**	**SVM**	**MP**	**KNN**	**ADA**	**JR**	**ARM**
Hand	Precision	0.46	0.33	0.42	0.53	0.55	0.67	0.47
	Recall	0.31	0.06	0.40	0.56	0.52	0.45	0.79
	F1-score	0.37	0.10	0.41	0.54	0.54	0.52	0.59
Urban	Precision	0.51	0.53	0.47	0.44	0.48	0.53	0.45
	Recall	0.41	0.25	0.45	0.50	0.45	0.44	0.79
	F1-score	0.46	0.34	0.46	0.47	0.46	0.48	0.57
Waterscape	Precision	0.58	0.65	0.53	0.54	0.55	0.62	0.50
	Recall	0.49	0.47	0.52	0.59	0.49	0.49	0.77
	F1-score	0.53	0.54	0.52	0.56	0.52	0.55	0.60
Crowd	Precision	0.58	0.65	0.53	0.45	0.50	0.52	0.41
	Recall	0.49	0.47	0.52	0.55	0.44	0.39	0.81
	F1-score	0.53	0.54	0.52	0.49	0.47	0.45	0.55
Face	Precision	0.60	0.68	0.55	0.48	0.53	0.61	0.47
	Recall	0.43	0.39	0.48	0.59	0.48	0.40	0.70
	F1-score	0.50	0.48	0.51	0.53	0.51	0.49	0.56
Person	Precision	0.54	0.61	0.48	0.46	0.49	0.54	0.43
	Recall	0.39	0.32	0.45	0.55	0.45	0.36	0.79
	F1-score	0.45	0.42	0.47	0.50	0.47	0.43	0.56
Walking	Precision	0.56	0.59	0.53	0.51	0.53	0.58	0.52
	Recall	0.52	0.57	0.52	0.59	0.52	0.59	0.80
	F1-score	0.54	0.57	0.52	0.55	0.53	0.58	0.63

Table 5 Average comparative performance for all concepts

AVERAGE	**DT**	**SVM**	**MP**	**KNN**	**ADA**	**JR**	**mean**
recall	0.39	0.48	0.32	0.25	0.31	0.36	0.35
F1-score	0.12	0.21	0.10	0.07	0.09	0.10	0.11

properties from both data mining and classification. In ARM, one of the research challenges is to develop a good interestingness measure for both rule generation and rule selection. In this paper, a novel video semantic concept detection framework that uses the ARM technique together with a new correlation-based interestingness measure is proposed. Our proposed framework first applies MCA to explore the correlation between 2-feature-value pairs and concept/non-concept classes, which is then applied to generate the candidate association rules. Next, the similarity information including the inter-similarity and intra-similarity values is calculated as a similarity-based score. The correlation information captured from MCA is reused and converted to a correlation-based score. These two scores are combined to calculate the final ranking scores to rank and select the final rule set for semantic concept detection (classification). Finally, the class label for each testing instance is determined by matching the selected association rules with

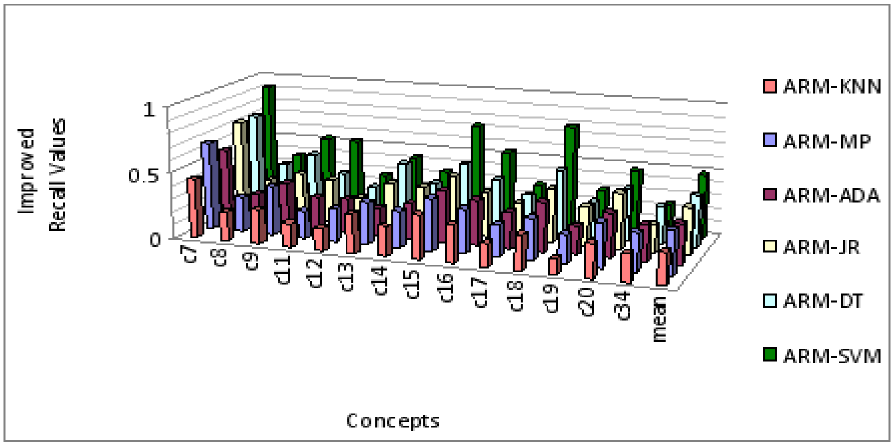

Figure 5 The Improved Recall Values for Each Concept and for Average

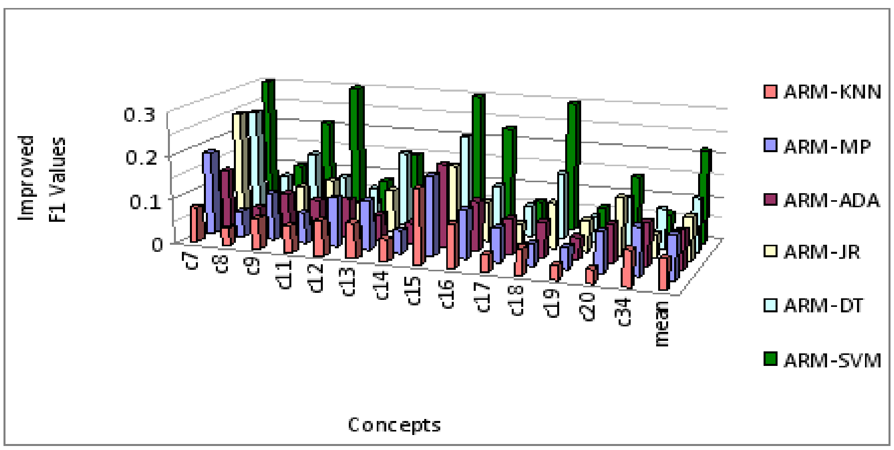

Figure 6 The Improved F1-score Values for Each Concept and for Average

the majority class. We have evaluated our proposed framework (ARM with a new correlation-based interestingness measure) by the detection performance of the videos taken from the TRECVID 2007 and 2008 projects. The experimental results show that our proposed framework demonstrates improved overall recall and $F1$-score performance over the other six classifiers that are commonly applied to concept detection. In the future, more researches on discretization will be done to investigate the effect of discretization to our framework and some new measurements will be applied. Moreover, the most important extension of this work is to automatically identify n-feature-value pairs where n is lager than 2 by utilizing the advantage of MCA. Then more discussions on n-feature-value pair rule generation and selection are necessary. Last, more videos will be processed so that more concepts can be evaluated to show the performance of our framework.

Acknowledgements

For Shu-Ching Chen, this research was supported in part by NSF HRD-0833093 and NSF OISE-0730065.

References

Agrawal, R. and Srikant, R. (1994) 'Fast algorithms for mining association rules in large databases', *Proc. International Conference on Very Large Data Bases (VLDB94)*, pp.487–499.

Fayyad, U. M. and Irani, K. B. (1992) 'On the handling of continuous-valued attributes in decision tree generation', *Machine Learning*, Vol. 8, No. 1, pp.87–102.

Fogarty, J., Tan, D., Kapoor, A. and Winder, S. (2008) 'CueFlik: interactive concept learning in image search', *Proc. ACM SIGCHI conference on Human factors in computing systems (CHI08)*, pp.29–38.

Geng, L. and Hamilton, H. J. (2006) 'Interestingness measures for data mining: A survey', *ACM Computing Surveys (CSUR)*, Vol. 38, No. 3, pp.9–es.

Lew, M. S., Sebe, N., Djeraba, C. and Jain, R. (2006) 'Content-based multimedia information retrieval: state of art and challenges', *ACM Tran. on Multimedia Computing, Communications and Applications*, Vol. 2, No. 1, pp.1–19.

Lin, L., Ravitz, G., Shyu, M.-L. and Chen, S.-C. (2007) 'Video semantic concept discovery using multimodal-based association classification', *Proc. IEEE International Conference on Multimedia and Expo (ICME07)*, pp.859–862.

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

Liu, B., Hsu, W. and Ma, Y. (1998) 'Integrating classification and association rule mining', *Proc. ACM International Conference on Knowledge Discovery and Data Mining (KDD98)*, pp.80–86.

Liu, B., Hsu, W., Chen, S. and Ma, Y. (2000) 'Analyzing the subjective interestingness of association rules', *IEEE Tran. on Intelligent Systems and Their Applications*, Vol. 15, No. 5, pp.47–55.

Liu, K.-H., Weng, M.-F., Tseng, C.-Y., Chuang, Y.-Y. and Chen, M.-S. (2008) 'Association and temporal rule mining for post-filtering of semantic concept detection in video', *IEEE Tran. on Multimedia*, Vol. 10, No. 2, pp.240–251.

Malik, H. H. and Kender, J. R. (2006) 'Clustering web images using association rules, interestingness measures, and hypergraph partitions', *Proc. ACM International Conference on Web Engineering (ICWE06)*, pp.48–55.

Nguyen, L., Tri, T., Huynh, H. X. and Guillet, F. (2008) 'Finding the most interesting association rules by aggregating objective interestingness measures', *Proc. Knowledge Acquisition: Approaches, Algorithms and Applications: Pacific Rim Knowledge Acquisition Workshop*, pp.40–49.

Rodda, S. and Shashi, M. (2007) 'A rough set based associative classifier', *IEEE International Conference on Computational Intelligence and Multimedia Applications*, pp.291–295.

Salkind, N. J. (2007) 'Encyclopedia of measurement and statistics', *SAGE Publications, Inc.*

Smeaton, A. F., Over, P. and Kraaij, W. (2006) 'Evaluation campaigns and TRECVid', *Proc. ACM International Workshop on Multimedia Information Retrieval (MIR06)*, pp.321–330.

Snoek, C. G. M. and Worring, M. (2008) 'Concept-based video retrieval', *Foundations and Trends in Information Retrieval*, Vol. 2, No. 4, pp.215-322.

Song, Q., Shepperd, M., Cartwright, M. and Mair, C. (2006) 'Software defect association mining and defect correction effort prediction', *IEEE Tran. on Software Engineering*, Vol. 32, No. 2, pp.69–82.

Tamir, R. and Singer, Y. (2006) 'On a confidence gain measure for association rule discovery and scoring', *The International Journal on Very Large Data Bases (JVLDB)*, Vol. 15, No. 1, pp.40–52.

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

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

Witten, I. H. and Frank, E. (2005) 'Data mining: Practical machine learning tools and techniques, Second Edition', *Morgan Kaufmann.*

Yang, G., Shimada, K., Mabu, S. and Hirasawa, K. (2008) 'A personalized association rule ranking method based on semantic similarity and evolutionary computation', *Proc. IEEE Congress on Evolutionary Computation (CEC08)*, pp.487–494.

Yu, B., Ma, W.-Y., Nahrstedt, K. and Zhang, H.-J. (2003) 'Video summarization based on user log enhanced link analysis', *Proc. ACM international conference on Multimedia (MM03)*, pp.382–391.

Zhao, N., Chen, S.-C. and Rubin, S. H. (2007) 'Automated multimedia systems training using association rule mining', *Proc. IEEE International Conference on Information Reuse and Integration (IRI07)*, pp.373–378.

Zhu, Y., Xiong, N., Park, J. H. and He, R. (2008) 'A web image retrieval re-ranking scheme with cross-modal association rules', *Proc. IEEE International Symposium on Ubiquitous Multimedia Computing (UMC08)*, pp.83–86.