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
l.lin2@umiami.edu, shyu@miami.edu

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

Abstract

The technique of performing classification using association rule mining (ARM) has been adopted to bridge the multimedia semantic gap between low-level features and high-level concepts of interest, taking advantages of both classification and association rule mining. One of the most important research approaches in ARM is to investigate the interestingness measure which plays a key role in association rule discovery stage and rule selection stage. In this paper, a new correlation-based interestingness measure that is used at both stages is proposed. The association rules are generated by a novel interestingness measure obtained from applying multiple correspondence analysis (MCA) to explore the correlation between two feature-value pairs and concept classes. Then the correlation-based interestingness measure is reused and aggregated with the inter-similarity and intra-similarity values to rank the final rule set for classification. Detecting the concepts from the benchmark data provided by the TRECVID project, we have shown that our proposed framework achieves higher accuracy than the classifiers that are commonly applied to multimedia retrieval.

Keywords: Interestingness Measure, Multiple Correspondence Analysis, Semantic Concept Detection.

1 Introduction

Motivated by a large number of requirements in multimedia retrieval applications, automatic video semantic concept detection has been rapidly developed. Some examples include the interesting event highlighter in sports videos, content-based image search engines, etc. Due to its high efficiency and good performance, classification using association rule mining (ARM) to perform video semantic concept detection has been studied [2][13].

There are two main phases in classification using ARM [4]. The first one is the phase of association rule discovery or rule generation, and the other one is the phase of association rule selection or rule ranking. The interestingness measure is one of the core elements to generate and/or rank the rules responding to the reduction of time and space costs in the mining process. In the rule discovery/generation phase, the interestingness measure could be used to prune the un-interesting feature-value pairs in order to improve the mining efficiency, such as the frequency count (minimum support) in the traditional ARM algorithm. In the rule selection/ranking phase, the interestingness measure could be used to score the rules so as to improve the mining accuracy, such as the evaluation criterion (minimum confidence) in the traditional ARM algorithm.

Typically, discovering association rules can be implemented by three categorized interestingness measures: objective measures, subjective measures, and semantic measures [1]. The objective measures are calculated based on probability, statistics, distance, or information theory. Most of the criteria depend only on the data, such as conciseness, coverage, reliability, peculiarity, and diversity. The subjective measures, such as surprisingness and novelty, consider both the data and the user's domain knowledge about the data. The semantic measures take into account the semantics and explanations of the feature-value pairs. In this category, utility and actionability consider the semantics of the data. In [3], we have introduced the utilization of *Multiple Correspondence Analysis (MCA)* as a semantic measure for association rule generation. MCA is a descriptive data analytic technique designed to analyze simple two-way and multi-way tables for more than two variables, containing some measure of correspondence between the rows and columns. By using MCA, the correspondence between the features and classes (columns) through the instances (rows) could be captured in a multimedia database. In this paper, we discover the 2-feature-value pair association rules generated by a similar approach that was applied to generate the 1-feature-value pair association rules in [3], but we focus on ranking those 2-feature-value pair association rules.

Ranking the association rules is a popular research topic and is considered as one of the most challenging aspects in classification using ARM [9]. There are various approaches to rank the association rules using the interestingness measures. Traditionally, the support and confidence values are mostly used in rule ranking. In [7], the length-first ranking strategy was used, that is, the longer the rules, the higher the priority of the rules. If two rules have the same length, the rules are ranked according to their confidence values. If two rules have the same confidence value, the rules are ranked according to their support values. A confidence gain measure for association rule scoring is introduced in [8], which takes advantages of both confidence measure and lift measure. The confidence gain presents the local confidence of a feature-value pair compared to its average confidence in the given database, and outperforms several association measures in the experiments. In [10], a novel conflict-based confidence measure with an interleaving ranking strategy for re-ranking association rules was proposed, which can better capture the inter-distance between a rule and a training data instance. The experiments achieved better performance than traditional confidence measures in both balanced and imbalanced data sets. The authors in [12] proposed a personalized association rule ranking method based on semantic similarity. The authors ranked the rules by using the similarity between the rules, the keywords assigned by user's interests, and the statistical information like support, confidence, and chi-square value. In [5], association mining 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 this paper, a novel video semantic concept detection framework facilitated with a correlation-based interestingness measure for association rule generation and selection is proposed. Our proposed framework consists of four stages. The first stage is to extract and normalize the low-level numerical features. Then the training data is first discretized and the the discretization results are used to discretize the testing data. Next, MCA is utilized to evaluate each of the extracted low-level features and identify the 2-feature-value pairs as the rules that better represent each one of the investigated concepts. In the last stage, 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 association rules for classification, where the concept class is determined by the majority class of the matched association rules. To evaluate our proposed framework, we use the concepts and data from TRECVID 2007 and 2008 [6], and make performance comparison 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 of 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 TRECVID datasets.

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

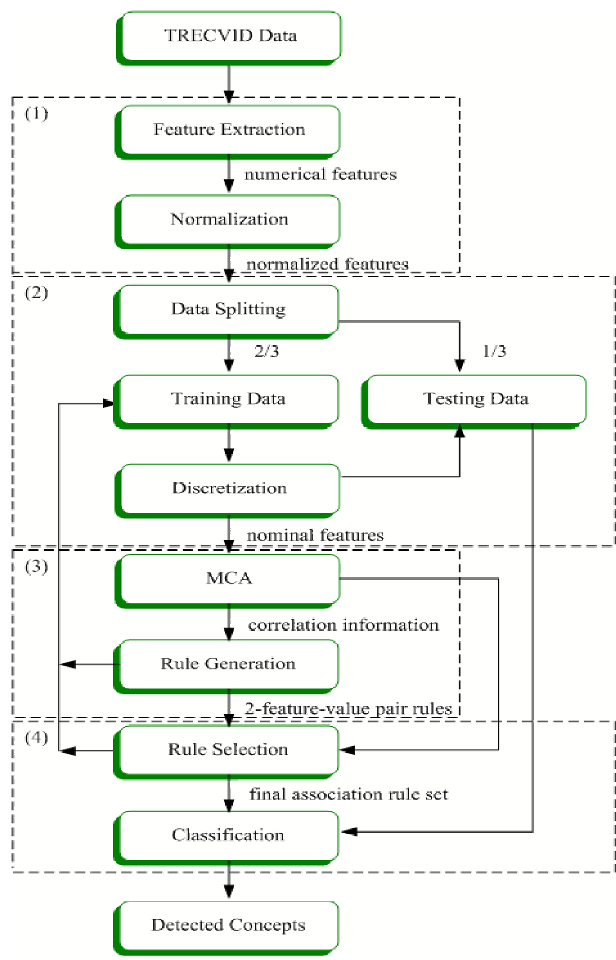

Figure 1. The Proposed Framework

2 The Proposed Video Semantic Concept Detection Framework

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 reuse of correlation-based interestingness measure for rule selection.

2.1 Framework Architecture

Our proposed framework is shown in Figure 1. It is composed of four stages as follows.

1. feature extraction and normalization;
2. data splitting and data discretization;
3. rule generation; and
4. rule selection and classification.

Since the shot boundary information has been provided by [6], the video shot boundary detection is beyond the scope of this paper. First, the audio-visual feature set based on the shot boundaries of each video is extracted. Totally, 28 continuous numerical features that were used in [2] are extracted and normalized. The data instances in a multimedia database are characterized by $F+1$ low-level features/columns, i.e., F numerical features A_f (where f=1 to F) and 1 nominal class label C_j (where j=1 to J and J is the total number of discrete classes) as shown in Table 1.

Table 1. Example data instances in the database after normalization

$feature_1$	$feature_2$	...	$feature_F$	$class_j$
0.23	0.78	...	0.05	C_1
0.17	0.67	...	0.02	C_4
0.10	0.59	...	0.21	C_5
...	...	...	...	...

Table 2. Example data instances in the database after discretization

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

For each concept, the data set includes all the data instances labeled with the target concept class (i.e., positive instances) and some randomly sampled data instances labeled with the non-concept class (i.e., negative instances). Then the data set is split into two parts, two-third of the data are used for training and one-third of the data are used for testing. Due to the fact that MCA requires the input data to be nominal, all the extracted features are discretized in this stage. The training data set is discretized into bins (partitions) which are then used to discretize the testing data set. These partitions generated by the discretization process are called *feature-value pairs* in our study.

After discretization, each feature A_f has K_f nominal feature-value pairs A_f^i (where i=1 to K_f). 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 partitions of the feature value ranges [0, 0.32865], (0.32865, 0.5044], and (0.5044, 1], respectively. Each feature would have several possible feature-value pairs and 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}, and $i2$=1 to K_{f2}. Some example data instances after discretization are shown in Table 2.

Assume that both rules $A_1^1 \wedge A_2^3 \Rightarrow C_p$ and $A_1^1 \wedge A_2^3 \Rightarrow C_n$ would be generated (as shown in Table 2). This indicates that the 2-feature-value pairs $\{A_1^1, A_2^3\}$ might represent both C_p and C_n, which is conflicting in classification. Hence, a rule generation strategy needs to be developed to address such an ambiguous situation. In [3], the utilization of MCA to analyze the multimedia data instances described by a set of low-level features and high-level concepts was explored. For rule generation, the correlation information between the 2-feature-value pairs ($\{A_{f1}^{i1}, A_{f2}^{i2}\}$) and classes ($C_p$ as the target concept class and C_n as the non-concept class) are calculated by applying MCA to the discretized training data set.

At the rule selection stage, the correlation-based interestingness measure that reuses the correlation information obtained from applying MCA is aggregated with the inter-similarity and intra-similarity values to calculate the rule ranking scores. 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. Finally, the selected 2-feature-value pair rules are used as the association rules for classification as follows. For each concept (class label), each testing data instance in the testing data set is checked to see if it includes any of the 2-feature-value pairs. For those 2-feature-value pairs that are found in the testing data instance, the majority class label of those matched 2-feature-value pairs is assigned to the testing data instance. If a tie occurs, which means that the number of matched positive rules is equal to the number of matched negative rules, then the data instance would be labeled as a positive instance. This procedure is repeated for each concept, i.e., using the selected 2-feature-value pair rules as the classification rules each time.

2.2 Rule Generation

Assume that there are N data instances in the multimedia database. The combination of each two feature-value pairs that do not belong to the same feature is considered as a 2-feature-value pair. Let the total number of such 2-feature-value pairs be S.

MCA codes the data by creating a binary column for each level with the constraint that one and only one of the columns gets the value 1 for each categorical (nominal) variable. This results in a matrix which is called the indicator matrix X with size $N \times S$. Usually, the inner product of the indicator matrix, called the Burt matrix $Y = X^T X$ (of size $S \times S$), is analyzed in MCA. Now, let the grand total of the Burt matrix be G and the probability matrix be $Z = Y/G$. The vector of the column totals of Z is a mass matrix M (of size $1 \times S$), and $D = diag(M)$. Let Δ be the diagonal matrix of the singular values, the columns of P be the left singular vectors (gene coefficient vectors), and the rows of Q^T be the right singular vectors (expression level vectors) in the SVD theorem. MCA will provide the principle components from singular value decomposition (SVD) given in Equation (1).

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

After projecting the multimedia data into a new space using the first and second principle components, the inner product of all possible 2-feature-value pairs and classes are calculated. The angles between the 2-feature-value pairs and classes are used as a measurement to represent the correlation. The 2-feature-value pairs whose angle values are smaller than an angle threshold are kept for rule discovery, where the angle threshold is determined by the one with the highest accuracy when applying the generated rule set with different thresholds to evaluate the training set. The pseudocode for generating a 2-feature-value pair rule is presented as follows.

RULE-GENERATION

$i \leftarrow 1;$
for $class_r \leftarrow class_p$ **to** $class_n$
  for $s \leftarrow 1$ **to** S
    $angle_s = cos^{-1} \frac{<pair_s, class_r>}{||pair_s|| * ||class_r||};$
    if $angle_s < threshold$ **then**
      $rule_i \leftarrow pair_s \Rightarrow class_r;$
    $i \leftarrow i + 1.$

Note that in order to reduce the computational cost, the same threshold is used for positive and negative association rules. Let $m1$, $m2$=1 to M ($m1 \neq m2$), $i1$=1 to K_{m1}, and $i2$=1 to K_{m2}. Equation (2) denotes the rules for class C, where C can be C_p (target concept class) or C_n (non-concept class).

$$A_{m1}^{i1} \wedge A_{m2}^{i2} \Rightarrow C. \tag{2}$$

2.3 Rule Selection

After the 2-feature-value pair rules as Equation (2) have been generated, the inter-similarity and intra-similarity are calculated to evaluate each rule.

For class C_j, assume that there are N_{jp} positive instances and N_{jn} negative instances, there are R_{kp} 2-feature-value pairs association rules generated for C_j, and there are R_{kn} 2-feature-value pairs association rules for non-C_j class. The intra-similarity and inter-similarity of the rules are defined as follows.

$$IntraSimilarity\_R_{kp} = \left(\sum_{counter1} 1 + \sum_{counter2} 0.25 + \sum_{counter3} 0\right) / N_{jp}. \tag{3}$$

$$InterSimilarity\_R_{kp} = \left(\sum_{counter1} 0 + \sum_{counter2} 0.75 + \sum_{counter3} 1\right) / N_{jn}. \tag{4}$$

$$IntraSimilarity\_R_{kn} = \left(\sum_{counter1} 1 + \sum_{counter2} 0.25 + \sum_{counter3} 0\right) / N_{jn}. \tag{5}$$

$$InterSimilarity\_R_{kn} = \left(\sum_{counter1} 0 + \sum_{counter2} 0.75 + \sum_{counter3} 1\right) / N_{jp}. \tag{6}$$

Here, $counter1$ 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, and $counter3$ is the counter that none of the feature-value pair is matched.

Please note that from our definitions, the larger the inter-similarity, the better the rule is, and the larger the intra-similarity, the better the rule is. Let $I_{kp} \in [0, 1]$ and $I_{kn} \in [0, 1]$. The harmonic mean of the inter-similarity and intra-similarity is calculated. The reason to use the harmonic mean is that compared to the arithmetic mean (simple average), the harmonic mean of a list of numbers tends strongly toward the least elements of the list. If the inter-similarity and intra-similarity values are large, the harmonic mean tends to mitigate the impact of large values. If the inter-similarity or intra-similarity value is small, the harmonic mean tends to aggravate the impact of small values.

$$\mathrm{I_{kp}} = \frac{2 \times InterSimilarity\_R_{kp} \times IntraSimilarity\_R_{kp}}{\left(InterSimilarity\_R_{kp} + IntraSimilarity\_R_{kp}\right)}. \quad (7)$$

$$\mathrm{I_{kn}} = \frac{2 \times InterSimilarity\_R_{kn} \times IntraSimilarity\_R_{kn}}{(InterSimilarity\_R_{kn} + IntraSimilarity\_R_{kn})}. \quad (8)$$

Next, the correlation information which is the angle values we got from the rule generation stage is reused. The angles are smaller than 90 degree, that is, either $angle_{kp}$ for positive class rules or $angle_{np}$ for negative class rules is smaller than 90 degree. It is obvious that the smaller the angle is, the more correlation it represents. Hence, the angles are converted to the forms using Equation (9) so that the larger the A_{kn} and A_{kp} ($\in [0,1]$) are, the better the rules are. Based on the same reason mentioned earlier, the harmonic means of I_{kp} and A_{kp}, and I_{kn} and A_{kn} are calculated. These values are used as the final ranking scores ($\in [0,1]$) for each 2-feature-value pair rule as given in Equation (10).

$$\begin{aligned} A_{kp} &= (1 - angle_{kp}/90); \\ A_{kn} &= (1 - angle_{kn}/90). \end{aligned} \quad (9)$$

$$score = \frac{2 \times I_{kp(n)} \times A_{kp(n)}}{\left(I_{kp(n)} + A_{kp(n)}\right)}. \quad (10)$$

Applying the generated rule set using different thresholds to evaluate the training set, the ranking threshold is the one which yields the highest accuracy. The 2-feature-value pairs rules whose ranking scores are larger than the threshold value are selected as the finial rule set.

3 Experiments and Results

Our proposed framework is validated using the videos taken from the video collection available to the TRECVID 2007 and 2008 high-level feature extraction task participants. The concepts used are two-people, building, street, sky, hand, urban, waterscape, face, person, and walking, whose descriptions can be found in [6].

The precision, recall, and F1-score metrics are adopted under the 3-fold cross-validation approach. In order to implement a convinced 3-fold cross-validation experiment, the positive instances and negative instances 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, it ensures that each data instance in the data set would be tested at least once.

To show the efficiency of our proposed framework, its performance is compared to those of the Decision Tree (DT), Support Vector Machine (SVM), and Neural Network

Table 3. Performance evaluation for five concepts

Two-people	**DT**	**SVM**	**NN**	**MCA**
Pre	0.51	0.00	0.44	0.35
Rec	0.19	0.00	0.25	0.94
F1	0.27	0.00	0.32	0.51
Building	**DT**	**SVM**	**NN**	**MCA**
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
Street	**DT**	**SVM**	**NN**	**MCA**
Pre	0.55	0.58	0.49	0.47
Rec	0.50	0.47	0.49	0.78
F1	0.52	0.52	0.49	0.59
Sky	**DT**	**SVM**	**NN**	**MCA**
Pre	0.62	0.64	0.56	0.46
Rec	0.48	0.45	0.51	0.81
F1	0.54	0.52	0.54	0.59
Hand	**DT**	**SVM**	**NN**	**MCA**
Pre	0.46	0.33	0.42	0.40
Rec	0.31	0.06	0.40	0.84
F1	0.37	0.10	0.41	0.54

Table 4. Performance evaluation for the rest concepts

Urban	**DT**	**SVM**	**NN**	**MCA**
Pre	0.51	0.53	0.47	0.45
Rec	0.41	0.25	0.45	0.77
F1	0.46	0.34	0.46	0.57
Waterscape	**DT**	**SVM**	**NN**	**MCA**
Pre	0.58	0.65	0.53	0.50
Rec	0.49	0.47	0.52	0.74
F1	0.53	0.54	0.52	0.59
Face	**DT**	**SVM**	**NN**	**MCA**
Pre	0.60	0.68	0.55	0.46
Rec	0.43	0.39	0.48	0.72
F1	0.50	0.48	0.51	0.56
Person	**DT**	**SVM**	**NN**	**MCA**
Pre	0.54	0.61	0.48	0.42
Rec	0.39	0.32	0.45	0.80
F1	0.45	0.42	0.47	0.55
Walking	**DT**	**SVM**	**NN**	**MCA**
Pre	0.56	0.59	0.53	0.51
Rec	0.52	0.57	0.52	0.80
F1	0.54	0.57	0.52	0.62

(NN) classifiers available in WEKA [11] using the same evaluation metrics and the same data sets. The WEKA project uses C4.5 algorithm for DT, Sequential Minimal Optimization (SMO) algorithm for SVM, and Multilayer Perceptron for NN. Note that the parameters of these three classifiers are set to the default values that WEKA has. The average precision (Pre), recall (Rec), and F1-score (F1) values obtained over the three folds are presented in Table 3 and Table 4, where columns 2 to 4 provide the performance of WEKA's DT, SVM, and NN, 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 concept detection framework achieves promising results, comparing to DT, SVM, and NN classifiers. It can also be observed that both the recall values and F1-scores of our proposed framework are higher than those of the three classifiers. The better recall values represent that more target concept instances are classified correctly. The better F1-scores mean that the recall values are increased without compromising the precision values too much. This observation demonstrates that our proposed framework using the new interestingness measure in both rule generation and rule selection can effectively improve the classification accuracy of the selected association rules.

4 Conclusion

This paper presents a novel video semantic concept detection framework that is facilitated with a new correlation-based interestingness measure and reuses it in both association rule generation and selection stages. The association rules are first generated by applying MCA to explore the correlation between 2-feature-value pairs and concept classes. Next, the correlation information is reused together with the inter-similarity and intra-similarity values to select the final rule set for semantic concept detection (classification). Finally, the target concept class is determined by matching the selected association rules with the majority class. Data taken from the TRECVID 2007 and 2008 video corpus is used to validate the detection performance of our proposed framework. The experimental results showed that our proposed framework demonstrates promising results with better overall recall and F1 performance over the DT, SVM, and NN classifiers that are commonly applied to multimedia retrieval.

5. Acknowledgment

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

References

[1] L. Geng and H. Hamilton. Interestingness measures for data mining: A survey. *ACM Computing Surveys (CSUR)*, 38(3):9–es, September 2006.

[2] 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*, pages 859–862, July 2007.

[3] 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)*, pages 316–321, December 2008.

[4] B. Liu, W. Hsu, and Y. Ma. Integrating classification and association rule mining. In *International Conference on Knowledge Discovery and Data Mining (KDD98)*, pages 80–86, August 1998.

[5] 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*, 10:240–251, February 2008.

[6] A. F. Smeaton, P. Over, and W. Kraaij. Evaluation campaigns and TRECVid. In *ACM International Workshop on Multimedia Information Retrieval (MIR06)*, pages 321–330, October 2006.

[7] Q. Song, M. Shepperd, M. Cartwright, and C. Mair. Software defect association mining and defect correction effort prediction. *IEEE Transactions on Software Engineering*, 32:69–82, February 2006.

[8] R. Tamir and Y. Singer. On a confidence gain measure for association rule discovery and scoring. *The International Journal on Very Large Data Bases (JVLDB)*, 15:40–52, January 2006.

[9] F. Thabtah. Challenges and interesting research directions in associative classification. In *IEEE International Workshop on Data Mining (ICDMW06)*, pages 785–792, December 2006.

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

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

[12] 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)*, pages 487–494, June 2008.

[13] N. Zhao, S.-C. Chen, and S. H. Rubin. Video semantic concept discovery using multimodal-based association classification. In *IEEE International Conference on Information Reuse and Integration (IRI07)*, pages 373–378, August 2007.