Leveraging Concept Association Network for Multimedia Rare Concept Mining and Retrieval

Tao Meng, Mei-Ling Shyu
Department of Electrical and Computer Engineering
University of Miami, Coral Gables, FL 33146, USA
Email: t.meng@umiami.edu, shyu@umiami.edu

***Abstract*—Automatic high-level semantic concept detection is a crucial step for multimedia data management, indexing, and retrieval. It is well-acknowledged that semantic gap poses a great challenge in multimedia content-based research. It becomes even more challenging when the concept of interest is extremely rare in the training data sets because of the poor modeling for the positive instances. In this paper, a Concept Association Network (CAN) is trained by selecting significant links to capture the strong associations among different concepts using association rule mining (ARM). By taking into account of the correlations and credibilities of reference concept nodes, the advantages of the reference nodes are utilized. Experimental results using TRECVID 2010 data sets show that by utilizing the proposed framework, the Mean Average Precision (MAP) values of all the concepts are improved, and the significant improvement of the MAP values of the rare concepts further attests the promising results.**

***Keywords*-content-based multimedia retrieval; concept association network; logistic regression; rare concept detection**

I. Introduction

The large amount of multimedia data such as videos on YouTube nowadays makes manual annotation infeasible. Therefore, automatically mining and annotating high-level concepts from multimedia data sets for the purposes of searching, indexing, and retrieving have become a popular research area [1][2][3]. The common approach used is to train a set of binary one-versus-all classifiers for each high-level concept based on the low-level features. The testing instance is tested and assigned a score by each classifier indicating the probability that the instance contains the corresponding concept. The main problem of this approach is semantic gap, which is the gap between low-level features and semantic concepts. This problem becomes even challenging when the positive instances in the training set are rare. On the other hand, the concepts do not occur independently in the multimedia data sets and they have some correlations. For example, the concepts 'airplane' and 'sky' have high chances to co-occur in the same image or video clip. Properly modeling the relationship among concepts and leveraging such information to improve the detection accuracy have received a large amount of attention.

Several frameworks have been proposed and they generally model the relationships in the following four ways: model vector, co-occurrence relationship, ontology, and probabilistic graphical network. The model vector consists of the scores output from the binary detectors and it is treated as a new feature vector. That new feature vector is then used for training and testing. The idea behind this approach is that the model vector provides a signature for one class so that the testing instance could be classified based on this signature [4]. One challenge of this approach is to overcome the error propagation issue because of the low accuracies of individual detectors. In terms of co-occurrence relationships, Chen et al. [5] utilized domain knowledge for the co-occurrence relationship to help the detection of the target concept using the information from the reference concept and gained 60% improvement for certain concepts in terms of MAP. More recent work integrates the information in linguistic ontology to model the correlations among different concepts so as to improve the detection rate [1][6]. One problem in this approach is its dependence on domain knowledge. Finally, some frameworks used the tree or graph structure with nodes representing concept models and edges representing association between concepts. For example, Choi et al. [7] modeled the contextual relationship in a probabilistic tree structure; while Aytar et al. [8] adopted a complete graph structure and used the conditional probabilities as weights for edges to combine the scores. One problem in this approach is the noisy inputs from the casual connections in the training data sets.

In data mining, association rule mining (ARM) is an important algorithm to identify the frequent patterns in a given data set. The frequent patterns represented by rules usually indicate the significant internal association among items. ARM has been widely utilized in multimedia concept detection frameworks [2][9]. Inspired by the idea of ARM, a Concept Association Network (CAN) is built by utilizing the Apriori algorithm [10] to capture the relationship among semantic concepts to improve the retrieval. Compared with different models introduced previously, the proposed CAN has four advantages. First, it captures only the significant connections among concepts so the noisy inputs from casual connections are eliminated. Second, the network is built using the labels given in the training data set so that there is no ground for the error propagation. Third, the link be-

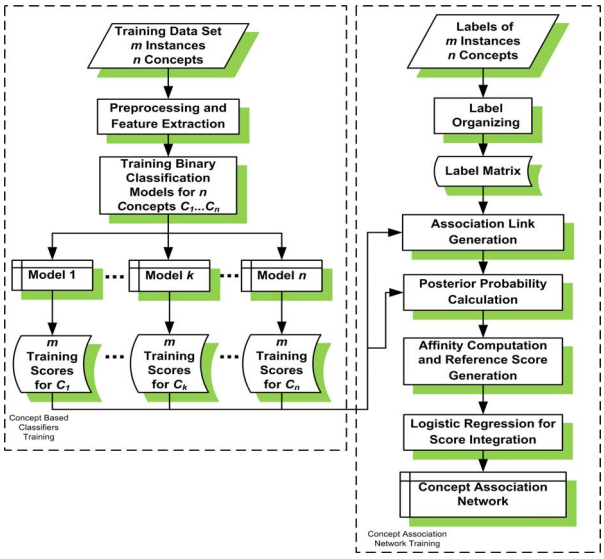

Figure 1. Training phase of the proposed framework

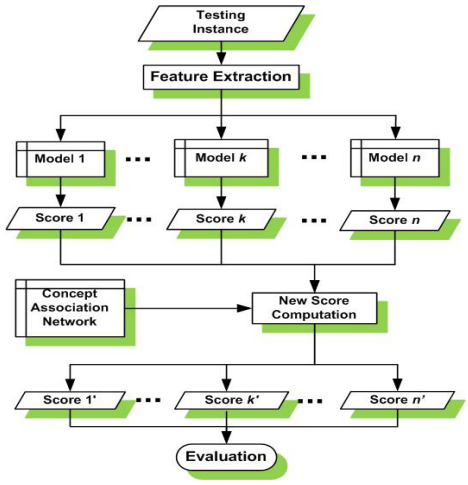

Figure 2. Testing phase of the proposed framework

tween nodes is modeled as bi-directional which matches the observations of the real world. Fourth, rather than capturing the relationship between only two nodes, the CAN could be easily extended to capture ternary or high order relationships among concept nodes by generating rules consisting of more than two items. Fifth, the construction of the CAN is fully automatic based on the training data, which shows the advantage over the probabilistic graphic model and ontology model since they usually rely on prior knowledge or domain knowledge.

The paper is organized as follows. In Section II, the proposed framework is introduced. Experimental results and discussions are shown in Section III. Section IV concludes the paper and describes future work.

II. The Proposed Framework

The proposed framework is shown in Figure I (training phase) and Figure II (testing phase). The training phase consists of the *Concept Based Classifiers Training Component* and the *Concept Association Network Training Component*. The former follows the architecture of the state-of-the-art multi-label multimedia concept mining system. Specifically, in a training data set, there are m instances (images or video shots) and n high-level concepts (such as *Outdoor* and *Sky*) to detect. The training instances are preprocessed and a set of features are extracted. Afterwards, n binary content-based classifiers (models) are trained for n concepts so that each model j $(1 \leq j \leq n)$ outputs m ranked scores for the j^{th} concept, represented by C_j in this paper. This component is included here for the purpose of completeness, and is not the main focus of the overall framework.

The *Concept Association Network Training Component* receives the ranked scores from the *Concept Based Classifiers Training Component* and mines the frequent itemsets in the label matrix (to be described in Section II-A) to build a CAN. In this component, several essential modules are included (to be introduced in details in the following subsections) to ensure that the association links retained in the CAN after rule pruning are significant and the related parameters are optimized. It is important to point out that this component does not depend on which models used in the *Concept Based Classifiers Training Component*.

In the testing phase, the same set of features as in the training phase are first extracted. For each testing instance, it receives one score from each content-based classifier. By leveraging the CAN, for a specific concept of interest, a new score which integrates the information from related concepts is generated as the final score. The evaluation criterion is Mean Average Precision (MAP).

Definition 1 (Concept-Class Pair): The **concept-class pair** is the representation of the class label for a certain concept. It is denoted as C_j^{ε}, where j $(1 \leq j \leq n)$ indicates the concept number and ε=1 or 0. When ε=1 (or ε=0), it is called positive (or negative) concept-class pair.

Definition 2 (τ-itemset): The **τ-itemset** is the set consisting of τ concept-class pairs, where $\tau = 1, 2, ..., n$. For example, $\{C_1^1, C_5^0\}$ is a 2-*itemset* (i.e., τ=2). Please note that C_j^1 and C_j^0 for the j^{th} concept cannot be in the same τ-*itemset*.

Definition 3 (Support): The **support** is defined as the number of occurrences of a τ-*itemset* in the training data set and is represented using $sup(\tau\text{-}itemset)$.

Definition 4 (Concept of Interest & Related Concept): The **Concept Of Interest (COI)** is defined as the concept to be detected (also called the target concept node in a CAN). The **Related Concept (RC)** is defined as the concept which has a significant correlation with COI (also called a reference concept node in a CAN).

Definition 5 (Self Score & Reference Score): The **self score** or **reference score** is defined as the score output from the content-based classifier corresponding to COI or RC, respectively.

A. Label Organizing

To mine the concept relationship from the training data set automatically, the labels of all concepts are organized in a label matrix. Assuming there are m instances and n classes in the training data set, the labels are organized in an $m \times n$ matrix $\boldsymbol{L} = \{l_{ij}\}$, $i = 1, 2, ..., m$ and $j = 1, 2, ..., n$, where $l_{ij} = C_j^1$ or $l_{ij} = C_j^0$ indicates the i^{th} instance is labeled as positive or negative for the j^{th} concept.

B. Association Link Generation

In order to utilize the relationship among different concepts, proper modeling of the links among those concepts is necessary. In general, the significant links should have the following characteristics. First, the links need to be significant compared to the other links. Second, the significance of the links depends on the *COI* to be detected. Take an example from the TRECVID 2010 [11] training data set. The importance of the link between C_6 (Animal) and C_{43} (Dogs) is different based on which concept to detect. In terms of C_6, the link between C_{43} and C_6 is very helpful because a correct positive label for dogs ensures a correct label for an animal. However, the link from C_6 to C_{43} is of less interest because an animal is not necessary a dog. Taking this into consideration, an ARM-based association link generation algorithm is proposed.

The algorithm works as follows. First, all 1-*itemsets* are generated for $\boldsymbol{L}$. Only the 1-*itemsets* $\{C_j^1\}$ consisting of positive concept-class pairs are retained. Second, all the candidate 2-*itemsets* are generated by combining the 1-*itemsets* with a minimum support of one. Afterwards, the candidate 2-*itemsets* which contain *COI* are organized together. Based on these 2-*itemsets*, a set of candidate rules which draw the conclusion that *COI* is positive are generated. In order to select the most significant rules, two rule pruning modules are incorporated into the framework.

Suppose that C_b is *COI* and C_a is *RC*, and one candidate rule is "$C_a^1 \rightarrow C_b^1$". The support-based rule pruning module addresses the significance of the rule from the perspective of *COI* and is modeled by the support ratio R_s defined in Equation (1).

$$R_s = \frac{sup(\{C_a^1, C_b^1\})}{sup(\{C_b^1\})}. \quad (1)$$

A threshold value $\alpha\%$ is used to select the rules with relatively high R_s values. Next, the interest-based rule pruning module is added to handle the significance from the *RC*'s perspective. The interest ratio R_i is defined in Equation (2). A threshold $\beta\%$ is used to select rules in a similar way as $\alpha\%$. The parameters $\alpha\%$ and $\beta\%$ are dynamically selected for each concept based on the training scores to maximize the average precision (AP) so that different preferences of the concept nodes are captured in the framework.

$$R_i = \frac{sup(\{C_a^1, C_b^1\})}{sup(\{C_b^1\}) \times sup(\{C_a^1\})}. \quad (2)$$

From the network point of view, if all the relationships among concepts are modeled in a network G=$\{V, L, W\}$, where V is a set of nodes (each node representing a concept), L represents a set of links, and each link has a weight in set W to model the relationship between two nodes. The selected significant rules could be viewed as the significant links from the *RC*s to the *COI*. These links are defined as the association links in the framework and form the core of a CAN.

C. Posterior Probability Calculation

Another important task is to model the credibility of the scores output from the content-based classifiers. In addition, instead of assigning a positive or negative label, modern multimedia retrieval systems usually output the ranking scores in a descending order to indicate the relevance of the retrieved results to a user query. Therefore, the traditional F_1 score, precision, and recall evaluation metrics are not suitable. In this study, a Bayesian posterior probability based score is used to integrate the information from the credibility of a model.

Assuming for an instance i, the detection score of C_j is $O(j,i)$. The output score $O'(j,i)$ for the C_j which encompasses the information of the credibility of the model is given in Equation (3).

$$O'(j,i) = \frac{p(O(j,i)|C_j = 1) \times p(C_j = 1)}{\sum_{z=0}^{1} p(O(j,i)|C_j = z) \times p(C_j = z)}, \quad (3)$$

where $p(C_j = 1)$ is the prior probability of C_j appeared in a data instance and is estimated by dividing the $sup(\{C_j^1\})$ by the total number of training instances. $p(C_j = 0)$ is one minus $p(C_j = 1)$ because there are only two possible cases. $p(O(j,i)|C_j = 1)$ is the conditional probability density function (pdf) $f_P(x) = p(x|C_j = 1)$ evaluated at $x = O(j,i)$, and $p(O(j,i)|C_j = 0)$ is the conditional pdf $f_N(x) = p(x|C_j = 0)$ evaluated at $x = O(j,i)$. To estimate $f_P(x)$ and $f_N(x)$, the Parzen-Window [12] approach is applied here. The kernel function used in this study is the standard normal distribution $\mathcal{N}(0,1)$, and the Parzen-Window estimation of the pdf for one dimensional random variable x is given by Equation (4).

$$p(x) = \frac{1}{v} \sum_{u=1}^{v} \frac{1}{\sqrt{2\pi}} \exp(-\frac{(x_u - x)^2}{2}). \quad (4)$$

Here x_u are the training instances and v is the total number of training instances. Using this equation, $f_P(x)$ and $f_N(x)$ could be estimated from the positive and negative training instances, respectively. In this way, not only does the information of the input score be included, but also the credibility of the model be integrated in the final score $O'(j,i)$.

D. Affinity Computation and Score Integration

After the association links are generated in Section II-B, the skeleton of the CAN is completed. The next step is to integrate the information from different sources. Specifically,

two questions need to be answered. The first question is how to analyze the link between the target concept node and reference concept nodes quantitatively. The second question is how to weigh the contributions of the self score and reference scores.

Different from most of the previous work, the link between two concept nodes is modeled as a bidirectional link in this study. This approach is adopted based on the observation of the asymmetrical relationship between semantic concepts in the real world so that the weight of the link depends on the direction. As far as the weight is concerned, inspired by the confidence measurement in ARM, we define the affinity of a link in Equation (5). Here, C_b is the target concept node, and $A_{a\to b}$ indicates the affinity of the link from C_a to C_b.

$$A_{a\to b} = \frac{sup(\{C_a^1, C_b^1\})}{sup(\{C_a^1\})}. \tag{5}$$

Hence, for C_b, the affinities of all the links from the reference concept nodes could be computed. Assume they form a set D. For one data instance i, the integrated information from all reference concept nodes is summarized in the integrated reference score $S_N(b,i)$, defined in Equation (6).

$$S_N(b,i) = \frac{\sum_{d\in D} A_{d\to b} \cdot O'(d,i)}{\sum_{d\in D} A_{d\to b}}, \tag{6}$$

where $O'(d,i)$ is computed for C_d and instance i using Equation (3).

While Equation (6) summarizes the reference scores, the information of the self score is still missing. It is important to weigh these two types of scores. In this study, the logistic regression algorithm is applied to determine the necessary weights. Specifically, for a concept C_b, let $\boldsymbol{x}^{(i)}$ be the column vector $[1\ S_N(b,i)\ O'(b,i)]^T$ and a parameter vector $\boldsymbol{\theta} = [\theta_0\ \theta_1\ \theta_2]^T$. The probability of an instance i to be positive is given by the logistic function in Equation (7).

$$g_{\boldsymbol{\theta}}(\boldsymbol{x}^{(i)}) = \frac{1}{1+\exp(-h_{\boldsymbol{\theta}}(\boldsymbol{x}^{(i)}))}; \tag{7}$$

$$h_{\boldsymbol{\theta}}(\boldsymbol{x}^{(i)}) = \boldsymbol{\theta}^T \boldsymbol{x}^{(i)}. \tag{8}$$

Here, $\boldsymbol{\theta}$ could be learned by minimizing the cost function in Equation (9) using the gradient descent algorithm. The updating rule for $\theta_q (0 \le q \le 2)$ is shown in Equation (10), where δ is the learning rate determined by empirical study based on the training data set, m is the number of training instances, and $y^{(i)}$ is either 1 or 0 indicating whether the instance is positive or negative. For a testing instance, the final score is computed using Equation (7) and the trained $\boldsymbol{\theta}$.

$$\begin{aligned} J(\boldsymbol{\theta}) &= -\frac{1}{m}[\sum_{i=1}^{m} y^{(i)} \log g_{\boldsymbol{\theta}}(\boldsymbol{x}^{(i)}) \\ &+ (1-y^{(i)}) \log(1 - g_{\boldsymbol{\theta}}(\boldsymbol{x}^{(i)}))]; \end{aligned} \tag{9}$$

$$\theta_q \leftarrow \theta_q - \delta \frac{\partial}{\partial \theta_q} J(\boldsymbol{\theta}). \tag{10}$$

III. Experimental Results

In this paper, the TRECVID 2010 [11] data set from the semantic indexing task is used for evaluation. It has 130 high-level concepts. After removing those keyframes which do not have any ground truth labels and data preprocessing, 118,581 keyframes from the training videos form the final training set and the corresponding labels are used to generate the label matrix. As mentioned in Section II, the *Concept Association Network Training Component* is able to adapt to any content-based classifiers (models). Therefore, CU-VIREO374 [13] is used and its detection scores are downloaded from [14] as the input. Since only the detection scores for the TRECVID 2010 testing keyframes were provided and the parameters of CAN need to be tuned using the training scores, the TRECVID 2010 testing keyframes are used in this paper. The labels for TRECVID 2010 were provided by NIST according to the collaborative annotation in 2011. In terms of the keyframes which contain only annotations for some concepts, the labels for the other concepts are treated as negative. Three-fold cross validation is adopted and three rounds of three-fold cross validation are performed.

To better evaluate the performance of the proposed framework, the approaches in [8] and [15] are implemented. In [8], a complete graph model was built without any link selection and the conditional probabilities were utilized to model the weights of the links. The parameter matrix w used in the implementation is computed using the least square method. In [15], the reference concepts were selected for each target concept in the ontology-based graphic model. Afterwards, the similarity between a reference concept and a target concept was computed based on the entropy values and used as the weight between the two concepts.

For a complete graph model under the current situation, there are 8385 links in total for 130 concepts if the bi-directional link between two concepts is counted as two individual links. After the association link generation, only about $3\% - 4\%$ of the links are retained and proved to be significant. For example, in one fold of the cross validation in the third round of the experiments, the retained links are 335 and each target concept node has 2.58 links on average, which indicates the selected links are sparse.

Table I shows the MAP values for all the 130 concepts for different numbers of retrieved instances. For example, "Top10" indicates the MAP value of the top 10 retrieved instances for all concepts, and the last column is the MAP value if all instances are retrieved. Each value in the table is the average of the three rounds of three-fold cross validation. "Baseline" corresponds to the MAP value of the raw scores, "Yusuf" indicates the approach in [8], "BH" indicates the method in [15], and "Proposed" indicates the proposed framework. "Impr.R1", "Impr.R2" and "Impr.R3" represent the relative improvement rates of the proposed framework compared with the "Baseline", 'Yusuf", and "BH", cor-

Table I
THE MAP VALUES OF 130 CONCEPTS FOR DIFFERENT NUMBERS OF RETRIEVED INSTANCES

Retrieved Instances	Top10	Top20	Top40	Top60	Top80	Top100	Top500	Top1000	Overall
Baseline	0.5218	0.4898	0.4481	0.4212	0.3999	0.3845	0.2807	0.2393	0.1382
Yusuf	0.4600	0.4304	0.4075	0.3925	0.3806	0.3693	0.2754	0.2374	0.1363
BH	0.5316	0.5011	0.4570	0.4308	0.4082	0.3927	0.2853	0.2440	0.1416
Proposed	0.5491	0.5143	0.4709	0.4428	0.4211	0.4051	0.2944	0.2515	0.1452
Impr.R1	**5.23%**	**5.00%**	**5.09%**	**5.13%**	**5.30%**	**5.36%**	**4.88%**	**5.10%**	**5.07%**
Impr.R2	**19.37%**	**19.49%**	**15.56%**	**12.82%**	**10.64%**	**9.69%**	**6.90%**	**5.94%**	**6.53%**
Impr.R3	**3.29%**	**2.63%**	**3.04%**	**2.79%**	**3.16%**	**3.16%**	**3.19%**	**3.07%**	**2.54%**
p_1	0.0033	0.0139	0.0022	0.0041	0.0025	0.0031	0.0011	0.0002	0.0002
p_2	0.0016	0.0034	0.0010	0.0033	0.0027	0.0038	0.0021	0.0017	0.0019
p_3	0.0089	0.0133	0.0030	0.0052	0.0037	0.0053	0.0029	0.0017	0.0016

Table II
THE OVERALL MAP OF RARE CONCEPTS

Rare Concepts	Top5Rare	Top10Rare	Top15Rare	Top20Rare	Top25Rare	Top30Rare
Mean Positive Number	4.2	10.6	15.87	20.8	25.96	30.6
Baseline	0.0010	0.0256	0.0245	0.0238	0.0215	0.0228
Yusuf	0.0003	0.0040	0.0053	0.0056	0.0067	0.0072
BH	0.0011	0.0286	0.0265	0.0254	0.0227	0.0239
Proposed	0.0012	0.0309	0.0289	0.0274	0.0245	0.0260
Impr.R1	**20.00%**	**20.70%**	**17.96%**	**15.13%**	**13.95%**	**14.04%**
Impr.R2	**300.00%**	**672.50%**	**445.28%**	**389.29%**	**265.67%**	**261.11%**
Impr.R3	**9.09%**	**8.04%**	**9.06%**	**7.87%**	**7.93%**	**8.79%**
p_1	0.01647	0.00455	0.00222	0.01243	0.01154	0.02795
p_2	0.00587	0.00882	0.00425	0.00379	0.00277	0.00177
p_3	0.01621	0.00348	0.01003	0.01276	0.00629	0.00502

respondingly. It could be seen that our proposed method outperforms "Baseline", "Yusuf", and "BH" in terms of the MAP values. In addition, in order to test the significance of the improvement, the Students' t test is applied. In this case, the null hypothesis is that the proposed framework does not lead to any improvement and the p-value indicates the maximum probability of making the Type I error. The p values for the corresponding MAP values are shown in the last three rows of Table I. "p_1", "p_2" and "p_3" denote the comparison of the proposed framework with baseline, "Yusuf" and "BH", correspondingly. The common criterion for p is 0.05 and therefore it shows that our proposed framework improves all MAP values significantly. The experimental results show that the "Yusuf" method is worse than the "Baseline". One possible reason is that the noisy inputs from the insignificant links actually confuse the models. Interestingly, the performance of the "Yusuf" method approaches baseline as the number of retrieved instances increases. This was also observed in [8] and one possible reason is that the scores of all models become smaller with the increase of the retrieved instances so the effect of integrating scores turns trivial. The "BH" approach performs better than the "Baseline" but still worse than the proposed framework. One possible reason is that their framework models the link between two concepts as a symmetric edge which does not match the asymmetrical relationship between semantic concepts in the real world.

In terms of rare concepts, the concepts are sorted in the ascending order according to the number of positive instances in the developing data set. Table II shows the MAP values for the top ranked rare concepts. For example, "Top5Rare" means the concepts which have top 5 fewest positive instances in the developing data set. "Mean Positive Number" means the mean value of the number of positive instances of the rare concepts in the corresponding column. It can be observed that the total number of instances in the developing set is 96516. Therefore, the top 5 rare concepts only have 4.2 positive instances for training. The following rows show the performance comparison in a similar way as in Table II. Here, it shows that our proposed framework improves the MAP values for the rare concepts consistently when compared with the "Baseline", "Yusuf", and "BH". The relative improvement of the MAP values for the rare concepts is higher than that of the whole 130 concepts. It indicates that the proposed framework is more effective for the rare concepts. The two possible reasons are as follows. First, compared with the classification model built on the target rare concept, the models of the reference concepts which have more training instances perform better, and the scores output from them are more reliable. By integrating such scores properly, the final output score for the rare concept is better. Second, generally speaking, the baseline MAP of the rare concepts is worse than the that of the overall 130 concepts. Therefore, the relative improvement of MAP for the rare concepts compared to that of all the concepts is larger. More future studies will be performed to shed light on the rationales in depth.

To further evaluate our proposed framework, a detailed analysis is performed to evaluate the contribution of each important module in it. The results are shown in Table III. The first column indicates the changing of the module. "No Change" denotes the original framework. "Remove Link

Table III
THE CONTRIBUTION OF EACH MODULE

Component Change	Overall MAP	Performance Drop
No Change	0.1452	0
Remove Link Selection	0.1245	0.0207
Remove Post. Proba. Calc.	0.1411	0.0041
Remove Logistic Regression	0.1435	0.0017
Remove Affinity Computation	0.1424	0.0028

Selection" indicates removing the link selection module to keep all the links between concepts. "Post. Proba. Calc." indicates using the raw scores directly without computing the posterior probabilities. "Remove Logistic Regression" indicates adding the self score and reference score directly without using the logistic regression algorithm to compute the proper weights. "Remove Affinity Computation" indicates the affinities in Equation (6) are set to 1 for all the reference concepts. The second column is MAP for all 130 concepts. The third column indicates the drop of the performance compared with the original framework. For example, if the "Link Selection" module is removed, the overall MAP drops to 0.1245 and the absolute value of drop is $0.1452 - 0.1245 = 0.0207$. It could be seen that the link selection module contributes the most to the performance gain. This observation again shows the importance of selecting the significant links. In addition, the posterior probability calculation module also makes relatively big contributions to the final performance gain.

IV. Conclusion and Future Work

This paper presents a framework that utilizes the concept relationships to improve the detection accuracies of the high-level concepts from multimedia data. A concept association network (CAN) is built based on the label matrix of the training data set and the generated association links. By integrating the information from the reference concept nodes properly, the final score is assigned to a testing instance for the target concept node as the output. Experimental results show that our proposed framework improves the MAP values for all the concepts significantly. Furthermore, the proposed framework provides even more promising results for the challenging rare concept detection problem. The future work includes extending the CAN with high-order relationships, refining the model of the affinities by integrating the tag information, developing more efficient algorithms for association link generation, and conducting more comparisons with other frameworks.

References

[1] L. Ballan, M. Bertini, A. D. Bimbo, and G. Serra, "Video annotation and retrieval using ontologies and rule learning," *IEEE Multimedia*, vol. 17, no. 4, pp. 80–88, Oct.-Dec. 2010.

[2] 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, no. 2, pp. 240–251, February 2008.

[3] M.-L. Shyu, Z. Xie, M. Chen, and S.-C. Chen, "Video semantic event/concept detection using a subspace-based multimedia data mining framework," *IEEE Transactions on Multimedia*, vol. 10, pp. 252–259, February 2008.

[4] J. R. Smith, M. Naphade, and A. Natsev, "Multimedia semantic indexing using model vectors," in *Proceedings of the 2003 International Conference on Multimedia and Expo*, July 2003, pp. 445–448.

[5] C. Chen, L. Lin, and M.-L. Shyu, "Utilization of co-occurrence relationships between semantic concepts in re-ranking for information retrieval," in *IEEE International Symposium on Multimedia*, December 2011, pp. 53–60.

[6] L. Bai, S. Lao, and J. Guo, "Video semantic concept detection using ontology," in *Proceedings of the Third International Conference on Internet Multimedia Computing and Service*, August 2011, pp. 158–163.

[7] M. J. Choi, A. Torralba, and A. S. Willsky, "A tree-based context model for object recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2011, to appear.

[8] Y. Aytar, O. B. Orhan, and M. Shah, "Improving semantic concept detection and retrieval using contextual estimates," in *Proceedings of the 2007 International Conference on Multimedia and Expo*, July 2007, pp. 536–539.

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

[10] R. Agrawal and R. Srikant, "Fast algorithms for mining association rules," in *Proceedings of the 20th International Conference on Very Large Data Bases*, September 1994, pp. 487–499.

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

[12] C. Archambeau, M. Valle, A. Assenza, and M. Verleysen, "Assessment of probability density estimation methods: Parzen window and finite Gaussian mixtures," in *Proceedings of the 2006 IEEE International Symposium on Circuits and Systems*, May 2006, pp. 3245–3248.

[13] Y.-G. Jiang, A. Yanagawa, S.-F. Chang, and C.-W. Ngo, "CU-VIREO374: Fusing Columbia374 and VIREO374 for large scale semantic concept detection," Columbia University, Tech. Rep., August 2008.

[14] Y.-G. Jiang, "Prediction scores on TRECVID 2010 data set," http://www.ee.columbia.edu/ln/dvmm/CU-VIREO374/, last accessed on September 8, 2011. [Online]. Available: http://www.ee.columbia.edu/ln/dvmm/CU-VIREO374/

[15] R. Benmokhtar and B. Huet, "An ontology-based evidential framework for video indexing using high-level multimodal fusion," *Multimedia Tools and Applications*, pp. 1–27, December 2011.