Enhancing Multimedia Semantic Concept Mining and Retrieval by Incorporating Negative Correlations

Tao Meng, Yang Liu, Mei-Ling Shyu, Yilin Yan
Department of Electrical and Computer Engineering
University of Miami
Coral Gables, FL 33124, USA
Email: {t.meng, y.liu39}@umiami.edu,
shyu@miami.edu, y.yan4@umiami.edu

Chi-Min Shu
National Yunlin University of Science and Technology
Department of Safety, Health, and
Environmental Engineering
Douliou, Yunlin 640, Taiwan, ROC
Email: shucm@yuntech.edu.tw

***Abstract*—In recent years, we have witnessed a deluge of multimedia data such as texts, images, and videos. However, the research of managing and retrieving these data efficiently is still in the development stage. The conventional tag-based searching approaches suffer from noisy or incomplete tag issues. As a result, the content-based multimedia data management framework has become increasingly popular. In this research direction, multimedia high-level semantic concept mining and retrieval is one of the fastest developing research topics requesting joint efforts from researchers in both data mining and multimedia domains. To solve this problem, one great challenge is to bridge the semantic gap which is the gap between high-level concepts and low-level features. Recently, positive inter-concept correlations have been utilized to capture the context of a concept to bridge the gap. However, negative correlations have rarely been studied because of the difficulty to mine and utilize them. In this paper, a concept mining and retrieval framework utilizing negative inter-concept correlations is proposed. Several research problems such as negative correlation selection, weight estimation, and score integration are addressed. Experimental results on TRECVID 2010 benchmark data set demonstrate that the proposed framework gives promising performance.**

***Key words*—Multimedia Semantic Mining and Retrieval, Negative Correlations, Information Integration.**

I. Introduction

Technologies for multimedia data are widely used in security, surveillance, activity monitoring, Web, mobile applications, medical imaging, disaster information management [1][2][3][4][5] and etc. With the rapid development of multimedia, communication and Web 2.0 technology, massive amounts of multimedia data have been increasingly available on desktops and smart mobile devices via Internet. Statistics shows that 72 hours of videos with all sorts of tags are uploaded to YouTube every minute and about 1.54 million photos are uploaded to Flickr every day [6]. Despite the ubiquity of multimedia data, effective management and retrieval of multimedia data are still considered challenging research topics. For example, the conventional tag-based indexing and searching technologies suffer from the noisy tag and missing tag issues. As a result, more and more researchers turn to content-based approaches [7][8][9][10][11][12][13][14][15]. One centric research task in content-based multimedia data retrieval field is multimedia concept mining and retrieval [16][17][18], which focuses on mining semantic concepts such as face, car, and airplane from the raw data directly. Accordingly, the annual TREC Video Retrieval (TRECVID) competition, organized by National Institute of Standards and Technology (NIST), has the "Semantic Indexing" task for concept detection from a large amount of videos collected from the Internet [19].

From a data mining point of view, the concept detection problem is essentially a multi-label classification problem. One video shot or image usually contains more than one concept. The conventional approach of solving this problem is the binary relevance approach [20]. This approach treats each concept as an individual class and converts one multi-label classification problem into multiple binary classification problems. Therefore, it ignores the correlations among different concepts. Nevertheless, the concepts are correlated in real-world multimedia data sets. For instance, some concepts co-occur more frequently, such as sky and cloud; while others rarely co-occur like ocean and road. Such types of correlations provide important context cues which assist concept detection.

There have been some previous studies in this research direction. Some researchers utilized the semantic model vectors to model correlations among concepts. In [21], the authors used the output scores from binary classification models to train another classifier, such as a K-nearest neighbor model, to generate new scores for each concept. Besides the model vector-based approaches, the graphical model is adopted to capture the inter-concept correlations due to its capability to model the relationships among multiple variables. In [22], the authors proposed a factor graph framework to set the spatio-temporal constraints among concepts. Meng et al. [23] carried out a series of studies to utilize the association affinity network to model positive inter-concept correlations and achieved improvements on the benchmark data sets. Another research direction is to rely on word ontology. For example, Ballan et al. [24] combined the linguistic ontology from WordNet [25] and association rule mining to help video retrieval. Similar studies are presented in [1][26][27][28].

Although previous work has addressed the utilization of inter-concept relationships from different perspectives, most work focused on utilizing the positive correlations among concepts [23][29][30]. The positive correlation in this scenario describes the case that the occurrence of one concept increases the chance of the existence of another one, such as the example of sky and cloud. Correspondingly, the negative correlation indicates that the existence of one concept decreases the probability of the occurrence of another one. One extreme example is "indoor" and "outdoor". Compared with their positive counterparts, negative correlations have rarely been studied. In [31], the authors claimed that the negative correlations contribute little in the performance improvement. However, negative correlations, which provide context cues from a different perspective, should be helpful intuitively.

In this paper, a framework which mines and utilizes negative correlations to enhance semantic concept detections is proposed. Given the fact that negative correlations are more challenging to be mined from multimedia data sets, a concept mining and retrieval framework with a two-step hierarchical negative correlation selection strategy is proposed to overcome this difficulty. In addition, instead of relying on the weights estimated from the labels only, the feature values are incorporated in estimating the weights using Multiple Correspondence Analysis (MCA). Finally, the score integration problem is formulated as a constrained optimization problem. The experimental result analysis indicates that negative correlations, if modeled properly, helps improve the concept detection accuracy.

This paper is organized as follows. In Section II, the proposed framework is introduced and important modules are explained in details. Section III presents our experimental results on the TRECVID 2010 data set and gives some insights. Section IV concludes the paper with a summary and identifies future research directions.

II. The Proposed Framework

Fig. 1(a) and Fig. 1(b) illustrate an overview of the proposed framework, which consists of a training section and a testing section. The training section consists of the "Multimedia Concept Mining" subcomponent and the "Concept Mining Enhancement" subcomponent. In the "Multimedia Concept Mining" subcomponent, for M data instances (e.g., images/video shots) and N concepts, N concept detection models are trained such that for each instance, the model k outputs a score measuring the likelihood that concept k exists in that data instance. This subcomponent is added here for the integrity of the framework, but it is not our main contribution. On the other hand, the "Concept Mining Enhancement" subcomponent is the main contribution of the proposed framework. First, all the class labels are organized into a label matrix so that each row contains the labels of different concepts for one data instance. Next, significant negative correlations are selected using this label matrix in the negative correlation selection module. A set of features are extracted from the original training data set to train the MCA-based negative weight estimation model. The weights generated from this model together with the output scores from the "Multimedia Concept Mining" subcomponent are normalized and used to train the regression-based score integration model. The selected negative correlations, MCA models, and regression models are stored for the testing section.

In the testing section, each testing data instance is plugged into all concept detection models to generate the testing scores. The same set of features used in the training section are extracted from that data instance to get the MCA-based weights. After the scores and weights are normalized, they are plugged into the regression-based score integration model to generate a new set of re-ranked scores. Finally, the new output scores are evaluated.

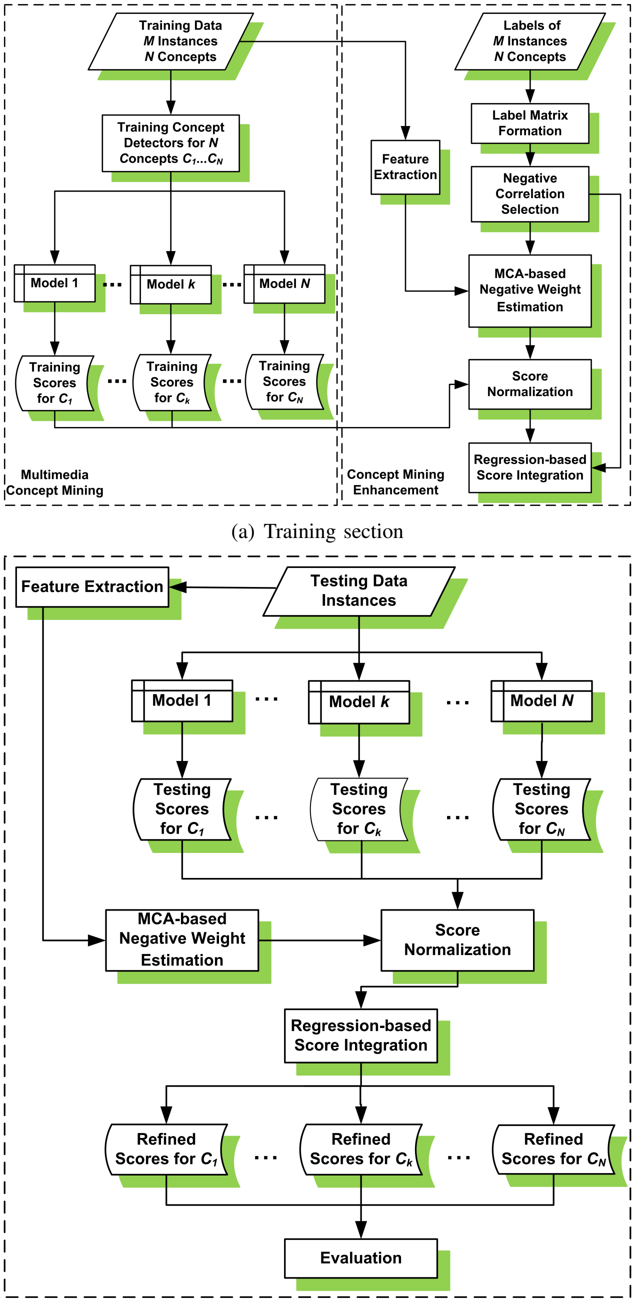
The details of each important module are introduced in the following subsections. For the sake of convenience, some commonly used terms in the proposed framework are explained here. A target concept is the concept to be detected, which is denoted as C_T. A reference concept is a concept that is negatively correlated with the target concept and is denoted as C_R. A data instance is positive if it contains a target concept and negative if it does not.

A. Negative Correlations Selection

The initial step of the overall framework is to select the significant negative correlations from the labels. While the co-occurrence of two concepts in one video shot/image increases the probability that they are positively correlated, the fact that one concept does not occur while the other appears does not indicate they are negatively correlated. For example, given an image which depicts a cat with no human face in the picture does not mean the appearance of a cat will exclude the appearance of a face. In this study, a two-step hierarchical negative correlation selection strategy is developed. First, a conditional probability-based coarse filtering approach is applied. The purpose of this step is to eliminate irrelevant correlations in an efficient way. Specifically, for a target concept C_T, C_T^+ and C_T^- represent the events that a data instance is positive or negative for C_T. Likewise, for a reference concept, C_R^+ and C_R^- represent the events that a data instance is positive or negative for C_R. If C_T and C_R are negatively correlated, the following conditions must hold.

$$\frac{P(C_T^-|C_R^+)}{P(C_T^-)} > 1; \tag{1}$$

$$\frac{P(C_T^+|C_R^-)}{P(C_T^+)} > 1. \tag{2}$$



(a) Training section

(b) Testing section

Fig. 1. The proposed framework

Here, $P(E)$ indicates the probability of an event E, where E is C_T^+, C_T^+, C_R^+, and C_R^-. The threshold values 1 in the above two inequalities are not selected arbitrarily, and they are necessary conditions for negative correlations. The first inequality indicates that the probability of C_T occurring decreases if C_R appears. On the other hand, the second inequality indicates that the probability of C_T occurring increases if C_R does not appear. From the association rule point of view, these two values are related to the conviction measurement introduced in [32]. Based on our empirical studies, this step eliminates 32.1% of possible correlations (from 16770 to 11556 concept pairs) and saves a great amount of computational complexity.

The second step is to filter the selected concepts more rigorously. The reason of adding this step is the observation that a huge number of data instances are not labeled and are given the inferred label 0. This is common in a large multimedia data set as manual labeling is very expensive. For instance, the concept pair "Indoor" and "Outdoor" should show perfect negative correlations. However, in our experiment, there are 104054 out of 115806 instances with negative labels for both concepts. Therefore, the conditional probability $P(C_{Outdoor}^+|C_{Indoor}^-)$ is only 0.0786, which severely deviates from 1. The problem could not be solved by discarding those data instances simply as it will introduce the bias that the two concepts are negatively correlated.

In order to tackle this difficulty, a novel strategy is proposed. The general assumption is that if two concepts are negatively correlated, their correlations would not be affected by the existence of the third concept, which is named as a control concept in this study. To formulate this problem, we define an integrated correlation factor (ICF) between the target concept and the reference concept, which is formulated in Equation (3).

$$ICF(T,R) = \frac{1}{|\Omega|-2} \sum_{D\in\Omega, D\neq T, D\neq R} \rho(C_T, C_R|C_D^+). \tag{3}$$

Here, Ω indicates the set of all concepts and $|\Omega|$ indicates the total number of concepts. C_D represents the control concept. C_D^+ is the condition that a data instance is positive for C_D, and $\rho(C_T, C_R|C_D^+)$ indicates the Pearson product-moment correlation coefficient [33] between the labels of C_T and C_R given C_D^+. The reasons for adding the control concept C_D are as follows. First, the data set provided by TRECVID has labels 1, 0, and -1 for a single data instance, and a 0 is given because no human annotator has watched this video [19]. Hence, a control concept, which selects data instances only if it is labeled as 1, would eliminate the 0-label cases. This increases the credibility of other labels for that data instance. Second, ICF represents an average quantitative metric of correlations under different conditions. For special cases where $\rho(C_T, C_R|C_D^+)$ is not defined, the default values are assigned as shown in Table I. In this table, "All C_t^1" indicates that all data instances are positive for C_t and "All C_t^0" indicates that all data instances are negative for C_t. Correspondingly, "All C_r^1" and "All C_r^0" have similar meanings. All the special cases happen when the data instances have unique labels for either C_t or C_r, in which case the Pearson correlation coefficients are not defined. As shown in this table, as long as C_t and C_r co-occur once, the value is set to a positive value, which imposes a relatively large penalty on that pair of concepts.

After sorting all the ICF values in an ascending order, we observe that the combined correlation coefficients follow a quasi-normal distribution. Hence, a Gaussian probability

TABLE I
THE VALUES TO BE SET UNDER SPECIAL CONDITIONS

C_t	C_r	Value To Set
All C_t^1	All C_r^1	1
All C_t^0	All C_r^0	0
All C_t^1	All C_r^0	the average value of the negative Pearson correlation coefficients
All C_t^0	All C_r^1	Same as above
Both C_t^0 and C_t^1 appear	All C_r^0	Same as above
All C_t^0	Both C_r^1 and C_r^0 appear	Same as above
All C_t^1	Both C_r^1 and C_r^0 appear	0.5
Both C_t^0 and C_t^1	All C_r^1	0.5

density function is a fit for all the ICF values. Different thresholds, depending on the significance levels such as 95% and 67%, could be set to select the significant negative correlations. As the Pearson product-moment correlation coefficients are symmetric, the concept pairs are selected. Therefore, the concept whose corresponding detector is less accurate is chosen as the target concept and the other one as the reference concept. The selection results are introduced in Section III-B.

B. MCA-based Negative Weight Estimation

The weights quantify the impact of a reference concept to a target concept. Some previous studies utilized the labels in the training data to estimate the weights, which did not consider the observed feature values for each data instance. In this paper, we treat the weights as a function of feature values. Formally, for a target concept C_T and a reference concept C_R, let F_i represent the visual features for data instance i, the probability that this data instance i is negative given F_i is represented as $P(C_T^-|F_i)$. Associating this probability with the reference concept, it could be expanded as follows.

$$\begin{aligned} P(C_T^-|F_i) &= P(C_T^-|C_R^+, F_i)P(C_R^+|F_i) \\ &+ P(C_T^-|C_R^-, F_i)P(C_R^-|F_i) \\ &= P(C_T^-, C_R^+|F_i) + P(C_T^-, C_R^-|F_i). \quad (4) \end{aligned}$$

This equation shows that the conditional probability $P(C_T^-|F_i)$ rely on both $P(C_T^-, C_R^+|F_i)$ and $P(C_T^-, C_R^-|F_i)$. From the correlation point of view, the summation of the two conditional probabilities quantifies the impacts of the reference concept to the target concept given the observed feature values. In order to estimate these probabilities, the MCA-based model is applied here. MCA is an extension of the standard correspondence analysis to more than two variables. It demonstrates the robustness and relatively high accuracy in modeling the posterior probability distribution [11][34]. Here, we use $P(C_T^-, C_R^+|F_i)$ as an example to show the algorithm that computes the weights. First, the training data instances which are negative for C_T and positive for C_R are selected and labeled as Type I instances. Second, other training data instances are labeled as Type II instances. Third, a MCA model is trained using the features, Type I labels, and Type II labels. Fourth, for each testing data instance, the features are plugged into the trained MCA model and a transaction weight is computed. This transaction weight represents the likelihood that the testing data instance belongs to Type I and is used to model the weights. The details of training the MCA model and calculating the transaction weights can be referred to [11]. $P(C_T^-, C_R^-|F_i)$ can be estimated in a similar way.

C. Score Normalization

The reasons of adding a normalization module are as follows. First, the transaction weights from the MCA-based weight estimation module and those from the original concept detectors need to be made compatible. Second, it is desirable that all scores are converted to well-calibrated probabilities. Given these requirements, the Bayes' conditional probability-based score normalization approach proposed in [23] is applied in our framework. As was tested in [23], this normalization approach shows two advantages. First, it provides a probabilistic estimation which meets the aforementioned two requirements. Second, it incorporates prior knowledge about the likelihood that the concept occurs in the training data. The details of this algorithm could be found in [23].

D. Regression-based Score Integration

After the scores are normalized, the next question is how to integrate the outputs from the target concept detectors, the reference concept detectors, and the MCA-based weight estimation models. Specifically, two scenarios, which correspond to a single reference concept and multiple reference concepts, need to be considered.

In the first case, a constrained optimization problem is formulated. Assume for one data instance i, the normalized scores from the target concept detector, the reference concept detector, and the MCA-based weight estimation module are represented by S_T^i, S_R^i, and S_M^i, respectively. Let $\boldsymbol{S}$ be the matrix such that each row $\boldsymbol{S}^i$ is a row vector $[1, S_T^i, S_R^i, S_M^i]$, and $\boldsymbol{\theta}$ is the column vector $[\theta_0, \theta_1, \theta_2, \theta_3]^T$. If the label of a data instance is represented by y^i, it indicates that it is positive ($y^i = 1$) or negative ($y^i = 0$). M is the total number of data instances. Assuming all training data instances are independent and identically distributed, a likelihood function is formulated as shown

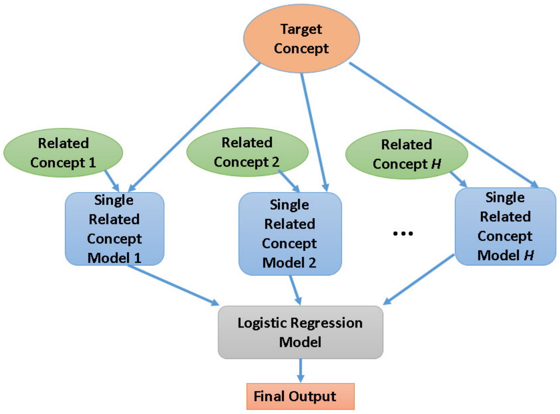

Target Concept
Related Concept 1
Related Concept 2
Related Concept *H*
Single Related Concept Model 1
Single Related Concept Model 2
...
Single Related Concept Model *H*
Logistic Regression Model
Final Output

Fig. 2. The Multiple Reference Concept Fusion Strategy

in Equation (5). Accordingly, a cost function is defined in Equation (6).

$$L(\boldsymbol{S};\boldsymbol{\theta}) = \prod_{i=1}^{M}(g(\boldsymbol{S}^i\boldsymbol{\theta}))^{y^i}\cdot(1-g(\boldsymbol{S}^i\boldsymbol{\theta}))^{(1-y^i)}, \tag{5}$$
$$\text{where } g(x) = \frac{1}{1+e^{-x}}.$$

$$J(\boldsymbol{S};\boldsymbol{\theta}) = -\log L(\boldsymbol{S};\boldsymbol{\theta}) + \lambda||\boldsymbol{\theta}||_2, \tag{6}$$
$$\text{subject to } \theta_1 \geq 0, \theta_2 \leq 0, \theta_3 \leq 0.$$

Here, λ is the regularization parameter to handle the overfitting problem, and $||\boldsymbol{\theta}||_2$ indicates the L_2 norm of $\boldsymbol{\theta}$. This is a constrained convex optimization problem and could be solved using the active set approach. The gradient of $\boldsymbol{\theta}$ is computed in Equation (7), where $\boldsymbol{S}$ has a dimension of M by 4, $\boldsymbol{y}$ is a column vector $[y^1, y^2, ..., y^M]^T$ which has a dimension of M by 1, $\boldsymbol{g}(\boldsymbol{U})$ indicates applying Equation (6) on each element in $\boldsymbol{U}$ and has a dimension of M by 1, and $\boldsymbol{I}$ is a 4 by 4 identity matrix.

$$\nabla_{\boldsymbol{\theta}} = \boldsymbol{S}^T(\boldsymbol{y} - \boldsymbol{g}(\boldsymbol{S\theta})) - 2\lambda\boldsymbol{I\theta}. \tag{7}$$

For a testing data instance f, the vector $\boldsymbol{S}^f$ can be generated in the same way as in the training section. Assume that the estimated parameter $\boldsymbol{\theta}$ is represented as $\hat{\boldsymbol{\theta}}$, the final output score S_F is computed as follows.

$$S_F = g(\boldsymbol{S}^f\hat{\boldsymbol{\theta}}). \tag{8}$$

In the second case, the multi-score collaboration approach is utilized. For each pair of the reference concept and the target concept, an integrated score is computed using Equation (8), and then the integrated scores are further fused using logistic regression. In this case, the score integration module could be viewed as a two-layer neural network. This process is illustrated in Fig. 2.

E. Evaluation

The proposed framework is evaluated from two perspectives. From the classification point of view, each concept detector is a binary classification model, and the area under precision recall curve (AUC) gives a comprehensive measurement of the classification performance. From the perspective of information retrieval, the users want to retrieve the most relevant data in the top-ranked results. Therefore, the ranking of the retrieved results matters. In order to capture the ranking information, the average precision (AP) value is a widely used metric in the multimedia concept retrieval domain. Specifically, for a given concept, ψ represents the number of the retrieved data instances, and G_n represents the total number of data instances containing that concept in the database. $Pre(a)$ indicates the precision of the a-th data instance in the ranking list. $Min(G_n, \psi)$ indicates the smaller value of G_n and ψ. The average precision at ψ (i.e., $AP@\psi$) is defined in Equation (9). The mean value of AP among all concepts is defined as the mean average precision (MAP) value.

$$\begin{aligned} AP@\psi &= \sum_{a=1}^{\psi}\frac{Pre(a)\times rel(a)}{Min(G_n,\psi)}, \qquad (9)\\ \text{where } rel(a) &= \begin{cases} 1, & \text{if instance } a \text{ is positive,}\\ 0, & \text{if instance } a \text{ is negative.}\end{cases}\end{aligned}$$

III. Experiments and Results

A. Data Sets

In this paper, the “IACC.1.A” testing data set from TRECVID2010 semantic indexing task [19] was used as the benchmark data set to compare all the different frameworks. This data set contains 200 hours of videos with the durations between 10 seconds and 3.5 minutes. The labels of 130 concepts were given by the collaborative annotation organized by NIST. After data pre-processing, there are 144774 video shots and one keyframe was extracted from each shot. The detection scores of all shots were downloaded from the DVMM Lab of Columbia University [35]. The list of concepts and the detailed explanations can be found in [19]. In order to increase the number of ground truth in the negative association selection module, TRECVID 2010 training labels are also utilized.

B. Negative Correlation Selection Results

For 130 concepts, all pair-wise associations are 8385. The top 10 selected associations from the conditional probability-based selection and the ICF-based selection are shown in Table II. For the conditional probability-based selection, the concept pairs are selected by adding the two probability ratios on the left sides in Equation (1) and Equation (2).

It can be seen that the proposed ICF-based selection approach selects more significant negative associations compared with the conditional probability-based approach. This indicates that the proposed ICF approach is effective. It should be pointed out that some negative correlations are caused by the definitions of concepts. For example, the “Two people” concept indicates that there must be

TABLE II
COMPARISON OF NEGATIVE ASSOCIATION SELECTION

Rank	Conditional Probability-based	ICF-based
1	Entertainment, Building	Road, Waterscape_Waterfront
2	Infants, Industrial	Indoor, Plant
3	Person, Helicopter_Hovering	Indoor, Vegetation
4	Person, Natural-Disaster	Daytime_Outdoor, Indoor
5	Person, Airplane_Flying	Indoor, Outdoor
6	Canoe, Bus	Suburban, Indoor
7	Telephones, Swimming	Indoor, Building
8	Cats, Person	Trees, Indoor
9	Canoe, Car_Racing	Male_Person, Female_Human_Face_Closeup
10	Person, Birds	Two_People, Single_Person

exactly two people in the video shot so the "Single_Person" concept does not occur, and the "Building" concept means the shots of an exterior of a building so it has the negative correlation with the "Indoor" concept. The full explanations of all concepts can be found in [19]. However, the conditional probability-based selection module is necessary from the computational point of view. Assume that the number of data instances is M and the number of concepts is N, the time complexity of the conditional probability-based selection module is $O(N^2M)$. Since there are no threshold tuning and each unit computation is a simple summation, this step is relatively efficient. However, the ICF-based approach has a complexity of $O(N^3M)$. In one round of the experiments (on MacBook Pro 2.6GHz Intel Core i7, 8GB RAM), it takes 12902 seconds to run all 8385 pair-wise associations for 234387 data instances using the ICF-based approach directly. However, the conditional probability-based selection module only takes 73.7 seconds. After running the conditional probability-based module, 2682 pairs are filtered, reducing the running time of ICF-based selection by 4126 seconds.

As introduced in Section II-A, we used the average value minus one standard deviation as the threshold, which corresponds to 0.67 significance level, to select the negative concept pairs. The selected correlations are shown in the right column of Table II. The seven target concepts used in the performance evaluation are "Road," "Indoor," "Daytime_Outdoor," "Suburban," "Trees," "Male_Person," and "Two_People".

C. Performance of the Proposed Framework

In order to evaluate the effectiveness of the proposed framework, it is compared with the following four frameworks. First, no modifications were made on the raw scores. Second, an intuitive solution which subtracts the scores of a reference concept from those of a target concept was applied. Third, a reference concept was selected randomly. Fourth, the domain adaptive semantic diffusion (DASD) framework [31] was applied. The MAP values at different numbers of the retrieved data instances, the precision recall curve, and AUC are reported. All results are the average of the three-fold cross validations over the seven selected concepts. Table III shows the MAP comparison results. The MAP values of the top 50 data instances are considered more important since users tend to focus more on the top-ranked results. The precision-recall curve and AUC are given in Fig. 3.

Some experiment details are given here. In the "Random Selection" framework, the randomly picked target concepts are: "Running" for "Road"; "Hand", "Old_People" and "Computer" for "Indoor"; "Beards" for "Daytime_Outdoor"; "Greeting" for "Suburban"; "Doorway" for "Trees"; "Chair" for "Male_Person"; and "Infants" for "Two_People". For the DASD framework, we kept all negative affinities as described in [31] and the number of iterations was set to 20, as we found that the number of iterations greater than 20 made the performance even worse.

The comparison of different frameworks gives the insights about the negative correlations. First, the intuitive solution of the "Subtraction" framework performs worse than the framework using the raw scores only. It indicates that the integration of negative correlations to enhance concept detection is a non-trivial research task. However, some improvements can be observed, such as "MAP@10", which indicates that negative correlations could be helpful if they are used properly. Second, the "Random Selection" framework decreases the accuracy tremendously, which indicates that the selection of negative correlations is important in our framework. Third, the "DASD" framework, which utilizes the graph diffusion algorithm, shows roughly the same performance as the framework using the raw scores. The results match to the ones given in the original paper [31], in which the authors claimed that the negative associations hardly improve the performance. On the other hand, our proposed framework gives the best performance in all frameworks. The possible reasons are two-fold. First, the proposed negative correlation selection module is able to capture the significant negative associations such as "Indoor" vs "Outdoor", given the challenging condition that lots of negative labels are inferred rather than manually annotated. Second, the MCA-based weight estimation model, which computes two conditional probabilities $P(C_T^-, C_R^+|F_i)$ and $P(C_T^-, C_R^-|F_i)$, is a better model than the one utilizing the probability estimated from labels only, such as in [31]. In conclusion, the experimental results show that the negative correlations could help concept detection

TABLE III
MAP VALUES AT DIFFERENT NUMBER OF INSTANCES RETRIEVED

Framework	MAP@10	MAP@20	MAP@30	MAP@40	MAP@50	MAP@100	MAP@200	MAP@500	MAP@2000
Raw	0.45077	0.40841	0.37285	0.35755	0.33353	0.24407	0.15928	0.13049	0.16538
Subtraction	0.47292	0.39973	0.36313	0.33913	0.31965	0.22274	0.15056	0.11547	0.12814
Random Selection	0.36005	0.31563	0.25743	0.23170	0.20680	0.13966	0.09827	0.08455	0.09293
DASD	0.48268	0.40204	0.36449	0.33404	0.32726	0.24311	0.15950	0.12220	0.13391
Proposed	0.86262	0.73554	0.65211	0.60537	0.57122	0.47294	0.36569	0.33970	0.40617

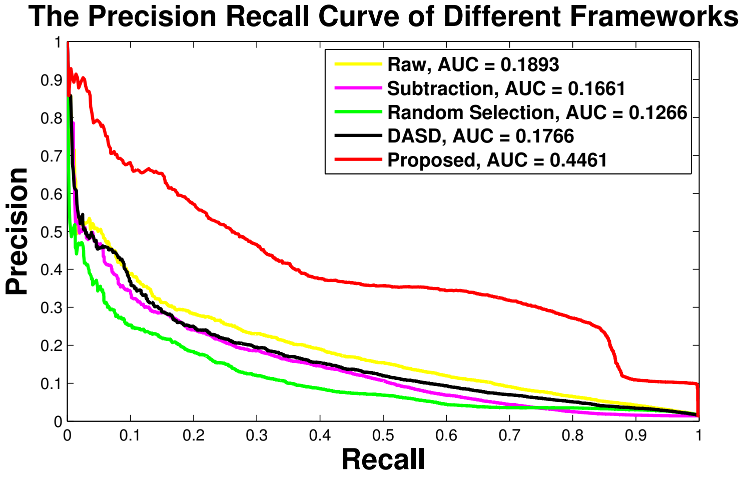


Fig. 3. The precision recall curves

if modeled properly.

IV. CONCLUSION AND FUTURE WORK

In conclusion, a framework that utilizes the inter-concept relationships is proposed to boost the accuracy of multimedia semantic concept detection. Different from most of the previous work, the proposed framework utilizes negative associations/correlations among concepts to enhance concept mining and retrieval. The negative correlation selection module, the MCA-based instance weight estimation module, and the score integration module form the core of the proposed framework. Experimental results demonstrate that the negative correlations between the reference concepts and the target concepts serve as the contextual cues for the target concept detection, while integrating this auxiliary information is a non-trivial task. By selecting and modeling the negative correlationsfig:PRCurve carefully, our proposed framework achieves promising results.

In the future, following this research direction, we will retain more negative correlations to see the relationship between the performance and the number of retained correlations. In addition, we will investigate the integration of positive and negative correlations. Last but not the least, high order correlations of different concepts will also be studied.

REFERENCES

[1] T. Meng and M.-L. Shyu, "Biological image temporal stage classification via multi-layer model collaboration," in *The 2013 IEEE International Symposium on Multimedia (ISM 2013)*, Anaheim, California, December 2013, pp. 30–37.

[2] M.-L. Shyu, T. Quirino, Z. Xie, S.-C. Chen, and L. Chang, "Network intrusion detection through adaptive sub-eigenspace modeling in multiagent systems," *ACM Transactions on Autonomous and Adaptive Systems*, vol. 2, pp. 9:1–9:37, 2007.

[3] M.-L. Shyu, C. Haruechaiyasak, and S.-C. Chen, "Category cluster discovery from distributed www directories," *Journal of Information Sciences*, vol. 155, pp. 181–197, 2003.

[4] Y. Yang, H.-Y. Ha, F. C. Fleites, and S.-C. Chen, "A multimedia semantic retrieval mobile system based on hidden coherent feature groups," *IEEE Multimedia*, vol. 21, pp. 36–46, 2014.

[5] L. Zheng, C. Shen, L. Tang, C. Zeng, T. Li, S. Luis, and S.-C. Chen, "Data mining meets the needs of disaster information management," *IEEE Transactions on Human-Machine Systems*, vol. 43, pp. 451–464, 2013.

[6] E. Gabarron, L. Fernandez-Luque, M. Armayones, and A. Y. Lau, "Identifying measures used for assessing quality of youtube videos with patient health information: A review of current literature," *Interactive Journal of Medical Research*, vol. 2, no. 1, March 2013.

[7] M. Chen, S.-C. Chen, and M.-L. Shyu, "Hierarchical temporal association mining for video event detection in video databases," in *IEEE International Workshop on Multimedia Databases and Data Management (MDDM07)*, April 2007, pp. 137–145.

[8] S.-C. Chen, M.-L. Shyu, C. Zhang, and M. Chen, "A multimodal data mining framework for soccer goal detection based on decision tree logic," *International Journal of Computer Applications in Technology*, vol. 27, pp. 312–323, 2006.

[9] S.-C. Chen, S. Sista, M.-L. Shyu, and R. L. Kashyap, "An indexing and searching structure for multimedia database systems," in *IS&T/SPIE conference on Storage and Retrieval for Media Databases 2000*, January 2000, pp. 262–270.

[10] X. Li, S.-C. Chen, M.-L. Shyu, and B. Furht, "An effective content-based visual image retrieval system," in *IEEE International Computer Software and Applications Conference (COMPSAC)*, August 2002, pp. 914–919.

[11] T. Meng and M.-L. Shyu, "Automatic annotation of drosophila developmental stages using association classification and information integration," in *The 12th IEEE International Conference on Information Resue and Integration (IRI 2011)*, Las Vegas, Nevada, August 2011, pp. 142–147.

[12] M.-L. Shyu, S.-C. Chen, M. Chen, and C. Zhang, “A unified framework for image database clustering and content-based retrieval,” in *the Second ACM International Workshop on Multimedia Databases (ACM MMDB’04)*, November 2003, pp. 19–27.

[13] M.-L. Shyu, S.-C. Chen, M. Chen, C. Zhang, and K. Sarinnapakorn, “Image database retrieval utilizing affinity relationships,” in *the First ACM International Workshop on Multimedia Databases (ACM MMDB’03)*, November 2003, pp. 78–85.

[14] C. Zhang, X. Chen, M. Chen, S.-C. Chen, and M.-L. Shyu, “A multiple instance learning approach for content based image retrieval using one-class support vector machine,” in *IEEE International on Multimedia & Expo*, July 2005, pp. 1142–1145.

[15] C. Zhang, S.-C. Chen, M.-L. Shyu, and S. Peeta, “Adaptive background learning for vehicle detection and spatio-temporal tracking,” in *the Fourth IEEE Pacific-Rim Conference On Multimedia*, December 2003, pp. 1–5.

[16] C. Chen, Q. Zhu, L. Lin, and M.-L. Shyu, “Web media semantic concept retrieval via tag removal and model fusion,” *ACM Transactions on Intelligent Systems and Technology*, vol. 4, pp. 61:1–61:22, 2013.

[17] D. Liu and M.-L. Shyu, “Semantic motion concept retrieval in non-static background utilizing spatial and temporal visual information,” *International Journal of Semantic Computing*, vol. 7, p. 43:67, 2013.

[18] 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.

[19] 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.

[20] E. A. Cherman, J. Metz, and M. C. Monard, “Incorporating label dependency into the binary relevance framework for multi-label classification,” *Expert Systems with Applications*, vol. 39, no. 2, pp. 1647–1655, February 2011.

[21] J. R. Smith, M. Naphade, and A. Natsev, “Multimedia semantic indexing using model vectors,” in *IEEE International Conference on Multimedia and Expo*, Baltimore, MD, June 2003, pp. 445–448.

[22] M. R. Naphade, I. Kozinetsey, T. S. Huang, and K. Ramchandran, “A factor graph framework for semantic indexing and retrieval in video,” in *the IEEE Workshop on Content-based Access of Image and Video Libraries*, Washington, DC, June 2000, pp. 35–39.

[23] T. Meng and M.-L. Shyu, “Leveraging concept association network for multimedia rare concept mining and retrieval,” in *IEEE International Conference on Multimedia and Expo*, Melbourne, Australia, July 2012.

[24] L. Ballan, M. Bertinti, A. D. Bimbo, and G. Serra, “Video annotation and retrieval using ontologies an rule learning,” *IEEE Multimedia*, vol. 17, no. 4, pp. 80–88, October-December 2010.

[25] G. A. Miller, “Wordnet: A lexical database for english,” *Communications of the ACM*, vol. 38, no. 11, pp. 39–41, 1995.

[26] X.-Y. Wei, C.-W. Ngo, and Y.-G. Jiang, “Selection of concept detectors for video search by ontology-enriched semantic spaces,” *IEEE Transactions on Multimedia*, vol. 10, no. 6, pp. 1085–1096, October 2008.

[27] T. Meng and M.-L. Shyu, “Model-driven collaboration and information integration for enhancing video semantic concept detection,” in *The 13th IEEE International Conference on Information Integration and Reuse (IRI2012)*, Las Vegas, Nevada, August 2012, pp. 144–151.

[28] C. Chen, T. Meng, and L. Lin, “A web-based multimedia retrieval system with mca-based filtering and subspace-based learning algorithms,” *International Journal of Multimedia Data Engineering and Management (IJMDEM)*, vol. 4, no. 2, pp. 13–45, 2013.

[29] L. Lin, M.-L. Shyu, and S.-C. Chen, “Association rule mining with a correlation-based interestingness measure for video semantic concept detection,” *International Journal of Information and Decision Sciences*, vol. 4, no. 2, pp. 199–216, 2012.

[30] Q. Zhu, L. Lin, M.-L. Shyu, and S.-C. Chen, “Effective supervised discretization for classification based on correlation maximization,” in *2011 IEEE International Conference on Information Reuse and Integration (IRI)*, 2011, pp. 390–395.

[31] Y.-G. Jiang, J. Wang, S.-F. Chang, and C.-W. Ngo, “Domain adaptive semantic diffusion for large scale context-based video annotation,” in *International Conference on Computer Vision (ICCV)*, Kyoto, Japan, September 2009, pp. 1420–1427.

[32] S. Brin, R. Motwani, J. D. Ullman, and S. Tsur, “Dynamic itemset counting and implication rules for market basket data,” in *1997 ACM SIGMOD international conference on management of data*, vol. 26, 1997, pp. 255–264.

[33] K. Pearson, “Notes on regression and inheritance in the case of two parents,” *Proceedings of the Royal Society of London*, vol. 58, pp. 240–242, 1895.

[34] 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*, December 2008, pp. 316–321.

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