Knowledge-Based Concept Score Fusion for Multimedia Retrieval

Manolis Falelakis, Lazaros Karydas, and Anastasios Delopoulos

Multimedia Understanding Group
Department of Electrical and Computer Engineering
Aristotle University of Thessaloniki, Greece
manf@mug.ee.auth.gr, kary@mug.ee.auth.gr, adelo@eng.auth.gr

Abstract. Automated detection of semantic concepts in multimedia documents has been attracting intensive research efforts over the last years. These efforts can be generally classified in two categories of methodologies: the ones that attempt to solve the problem using discriminative methods (classifiers) and those that build knowledge-based models, as driven by the W3C consortium. This paper proposes a methodology that tries to combine both approaches for multimedia retrieval. Our main contribution is the adoption of a formal model for defining concepts using logic and the incorporation of the output of concept classifiers to the computation of annotation scores. Our method does not require the computationally intensive training of new classifiers for the concepts defined. Instead, it employs a knowledge-based mechanism to combine the output score of existing classifiers and can be used for either detecting new concepts or enhancing the accuracy of existing detectors. Optimization procedures are employed to adapt the concept definitions to the multimedia corpus in hand, further improving the attained accuracy. Experiments using the TRECVID2005 video collection demonstrate promising results.

1 Introduction

The exponential growth of multimedia content during the last decade has made efficient indexing and retrieval a necessity. The appearance of evaluation frameworks such as TRECVID [1] and baseline frameworks such as Mediamill [2] and Columbia374 [3] reflects this trend.

Much of the research effort towards this direction is governed by the development of discriminative concept classifiers, often yielding satisfactory results. These can be further improved by using classifier score fusion [4] with concepts either selected manually [5] or determined automatically [6,7].

On the other hand lie methods relying on knowledge. These are mainly based on inference using expressive Description Logics [8]. Extensions of these methods that model certainty using Fuzzy Description Logics [9] have recently been successfully employed for multimedia retrieval [10,11]. These, however, have certain restrictions, imposed mainly by the computational cost of reasoning which can become prohibitive when dealing with large concept collections.

J. Liu et al. (Eds.): AMT 2009, LNCS 5820, pp. 126–135, 2009.

Present paper proposes a methodology that aims at combining virtues from both aforementioned approaches.The output scores of concept classifiers are fuzzified and used by formal fuzzy knowledge models to detect semantic concepts in multimedia. This way, the computationally intensive training of new classifiers is unnecessary. Furthermore, based on a notion that (existential or universal) quantifiers are of no use in this particular case, the knowledge models adopted are very simple, minimizing the complexity of reasoning.

The target of our methodology is two-fold, aiming at (i) providing an inexpensive yet reliable means of extending existing concept detector schemes and (ii) enhancing the detection accuracy of concepts for which specialized classifiers already exist. A virtue of this approach is that it does not dictate the type of features used. In fact, concept classifiers used may be trained using completely different feature vectors.

Moreover, our method is coupled with optimization schemes that help to adapt the fuzzy definitions to the dataset in-hand. To this end, we use a genetic algorithm which is accompanied with of k-fold cross validation [12] and RankBoost [13] resampling algorithms, in order to avoid over-fitting on the training sets and provide a more modest estimation on the behavior when applied to the test set.

We must point out that our method does not necessarily provide better results than concept classifiers in all cases. However, exploitation of the estimate of its performance can help us determine when its use is of potential benefit for classifier enhancement.

Experiments conducted using the LSCOM concept ontology on a TRECVID dataset demonstrate the effectiveness of the proposed approach on real data. Results show that new concepts can be efficiently defined, often attaining performance comparable to specifically trained concept classifiers (but with minimal computational effort) while it can provide significant improvement to detection of corresponding concepts for which classifiers exist.

The next section is devoted to describing the Fuzzy Models, while section 3 describes the fuzzy degrees adaptation procedure. Section 4 presents experiments conducted on real data and section 5 concludes the paper and includes some future directions.

2 Fuzzy Definition Models

The main idea of our approach is to rely on the result of reliable concept-classifiers to infer on other concepts.

Concept classifiers treat an image as a whole and provide information whether it (up to a certain degree) belongs to a certain class or not. However, they provide no evidence on the existence and the type of possible interrelations between the detected concepts. Due to this reason, it makes no sense to model these relations using object properties ('roles' in the Description Logics terminology).

Based on the previous notions we adopt a language which can contain statements based on conjunction, disjunction and negation operators, i.e., disregarding quantifiers and other expressivity tools provided by Description Logics. More formally, the expressions are constructed according to the following syntax rule:

$$\begin{array}{lll}
C, D & \longrightarrow & A \mid \qquad \text{(atomic concept)} \\
 & & \top \mid \qquad \text{(universal concept)} \\
 & & \bot \mid \qquad \text{(bottom concept)} \\
 & & \neg C \mid \qquad \text{(negation)} \\
 & & C \sqcup D \mid \qquad \text{(union)} \\
 & & C \sqcap D \qquad \text{(intersection)}
\end{array}$$

Furthermore, we allow subsumptions to hold up to a certain degree, i.e., we model uncertainty in a way similar to the one of [9]. In this direction, a concept S_i is subsumed by a concept C_i to the degree f_i, as displayed in equation 1.

$$< S_i \sqsubseteq C_i, f_i > \tag{1}$$

Let a hierarchy $\mathcal{T}$ of such subsumptions, according to which, concept C subsumes concepts $S_1 \ldots S_k$, i.e.,

$$\mathcal{T} = \begin{cases} < S_1 \sqsubseteq \mathcal{C}, f_1 >, \\ < S_2 \sqsubseteq \mathcal{C}, f_2 >, \\ \qquad \ldots, \\ < S_k \sqsubseteq \mathcal{C}, f_k > \end{cases} \tag{2}$$

Inference on the degree of the existence $\mu(C)$ of concept C, based on the existence of concepts S_i as given by the fuzzified classifier output $\mu_c(S_i)$, is made according to the *type 1* definition which is of the following form

$$\mu(C) = \mathop{\mathcal{U}}_{i}(\mathcal{I}(\mu_c(S_i), f_i)) \tag{3}$$

where the operators $\mathcal{U}$ and $\mathcal{I}$ denote fuzzy union and intersection operators respectively.[1]

The existence of the other concepts of $\mathcal{T}$ is computed with definitions of *type 2* that take the following form

$$\mu(S_i) = \mathcal{I}(\mu_c(C), \mathop{\mathcal{I}}_{j \neq i}(\mathcal{N}(\mathcal{I}(\mu_c(S_j), f_j)))) \tag{4}$$

where the operator $\mathcal{N}$ denotes fuzzy complement (negation).

To illustrate these with an example, consider the hierarchy depicted in figure 1 that can be encoded as

$$\mathcal{T} = \begin{cases} < Car \sqsubseteq Vehicle, f_{Car} >, \\ < Bus \sqsubseteq Vehicle, f_{Bus} >, \\ < Motorcycle \sqsubseteq Vehicle, f_{Motor} > \end{cases} \tag{5}$$

Definitions of type 1, computed with equation 3 suggest that we compute the degree of existence of ’Vehicle’ as the logical union of the degrees of existence of ’Car’, ’Bus’ and ’Motorcycle’, meaning that a ’Vehicle’ is *a ’Car’ or a ’Bus’ or a ’Motorcycle’*.

[1] The equation can be written this way (with union taking multiple inputs) due to the associativity and commutativity properties of fuzzy norms.

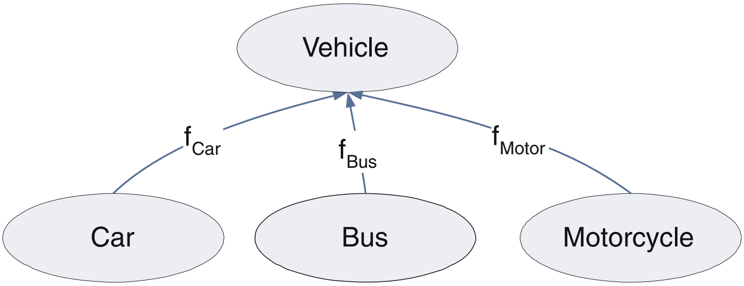

Fig. 1. An example of a simple hierarchy

On the other hand, with a type 2 definition, we can compute the existence of a 'Car', defining it as *a 'Vehicle' and not a 'Bus' and not a 'Motorcycle'*, when scores for the latter concepts exist.

Note that by forming definitions of type 2, we make a disjointness assumption, i.e., all hierarchy siblings are assumed to be disjoint. This could not always be the case. For instance an image may contain two sibling concepts (such as a 'Car' and a 'Motorcycle in our example) at the same time. However, this assumption leads us to easy definition extraction and proves to work in practice as shown in the experiments of section 4.

As stated before, our approach can also become useful when the classifier score for the concept under examination is available and the goal is to improve the retrieval performance. In this case, in order to compute $\mu(C)$ given the fuzzified output $\mu_c(C)$ of the corresponding classifier, we use $\mu_c(C)$ and eq. 3 in a disjunctive manner, and a type 1 definition takes the following form

$$\mu(C) = \mathcal{U}(\mathcal{I}(\mu_c(C), f_C), \underset{i}{\mathcal{U}}(\mathcal{I}(\mu_c(S_i), f_i))) \tag{6}$$

In a similar manner derives inference for concepts of type 2:

$$\mu(S_i) = \mathcal{U}(\mathcal{I}(\mu_c(S_i), f_i), \mathcal{I}(\mu_c(C), \underset{j \neq i}{\mathcal{I}}(\mathcal{N}(\mathcal{I}(\mu_c(S_j), f_j))))) \tag{7}$$

Subsumptions in the form of Eq. 1 can be extracted from crisp domain ontologies such as LSCOM [14], and use optimization techniques to compute the degrees f_i, fuzzifying the hierarchy (i.e., making subsumptions hold up to a certain degree) and adapting to the dataset under examination.

3 Adaptation to the Corpus In-Hand

In order to fuzzify the knowledge source we have to compute the weights (i.e. the values for f_i) presented in the previous section. This is essentially an optimization problem.

The target is to compute f_i with respect to the current dataset with the goal of maximizing the average precision for each of the defined concepts. For the parametric norms (we have experimented with Dubois-Prade and Yager class) the value of the norm parameter is also determined by training.

As the optimization surface proves to contain local minima we employ a genetic algorithm for this task. The fitness function to be minimized is the arithmetic negation of the average precision.

In order to improve the generalization properties of the computed weights we use two resampling methods; k-fold cross validation and a modified version of RankBoost[13].

3.1 *k*-Fold Cross Validation

In k-fold cross validation the original training samples are firstly partitioned into k subsets, called folds. Due to the nature of the video dataset (scarce positive samples tend to appear in bursts) the partitioning of the training set into folds is made by evenly distributing the positive samples in every fold. This assures that all k folds are representative of the whole set, with respect to the prior probability of positives.

Then the genetic algorithm is called to train the fuzzy weights using as training set all the samples in $k-1$ folds, leaving one fold out, which is used for validation. This procedure is repeated k times until each fold has been used for training $k-1$ times and for validation exactly once.

Finally, we end up with k sets of fuzzy weights which are averaged to obtain a single set. These are the values for f_i that are then incorporated to the inference function.

The results obtained by this method in the training set are used as a modest estimate of the expected average precision of a concept in the test set, therefore providing a way to determine the potential improvement of performance in classifier enhancement. Since the method is not so prone to over-fitting on the training set the selection of the potentially improved classifiers can be made by setting a performance threshold.

3.2 RankBoost

Boosting is a technique used in machine learning that tries to create a set of weak learners and combine them into a single strong learner.

The optimization scheme used here is a modified version of the RankBoost algorthm (see [13]). The algorithm runs in T rounds, randomly choosing all the positive and an approximately equal number of negative samples to form a new training set for the next round. A distribution function is initialized for the purpose of determining which samples are going to be used for training in each round. This distribution function is updated in each round, so as to promote the selection of samples that were misclassified in the previous rounds.

In each round, the genetic algorithm is called to train the weights with respect to the subset of the samples that have been chosen as training set. This means that the genetic algorithm training procedure will rerun from scratch exactly T times.

Finally, RankBoost computes T sets of fuzzy weights and inference is performed using a combination of them, which is a weighted average with the accuracy obtained in each round.

4 Experiments

For the purpose of our experiments we have used the Columbia374 [3] set of semantic concept detectors, which also includes the ground truth, the features, and the results of the detectors over the TRECVID datasets.

Our dataset consisted of the 47 videos of the TRECVID2005 development set that were not used for training the Columbia classifiers. These videos (corresponding to 20054 shots) were split to a training (23 videos) and an evaluation (24 videos) set, each one containing about 10000 shots.

In order to form the definitions automatically, we used a cut-down version of the LSCOM, that includes the 374 concepts of our dataset and exploited its hierarchy. We have conducted two experiments for evaluating our method. In the first experiment we demonstrate how new concepts can be defined, in the presence of no adequate classifier, while in the second one we perform a type of query expansion in order to improve the accuracy of concept detection.

4.1 Concept Scalability

This experiment simulates the case of extending the vocabulary of a multimedia retrieval system using knowledge. This is fully scalable and extensible as new concepts can be defined recursively, while training requires minimal effort compared to building a classifier model for every new concept.

The definitions used for this experiment are of the form displayed in equations 3 and 4.

We have chosen to define concepts, already existing in the ontology but without taking into account the actual classifier scores for them during inference. Instead, we use these scores as a baseline for comparison purposes.

Figure 2 displays the attained average precision for several concepts, using this kind of definitions. The concepts here were selected based on a certain threshold imposed on their performance on the training set when using the cross validation method. This gave us a hint of their performance on the evaluation set.

Commenting on figure 2, our methodology seems to yield very satisfactory results, often comparable to the ones of specifically trained classifiers. In some cases (see 'Road Overpass' for example), it outperforms the corresponding classifier. This is very important considering the computational cost of training the latter. Finally, in every case, the use of fuzzy weights, adapted to the set in-hand, significantly improves the performance.

4.2 Classifier Enhancement

In this experiment classifier scores are taken into account and the definitions formed correspond to the ones of the equations 6 and 7.

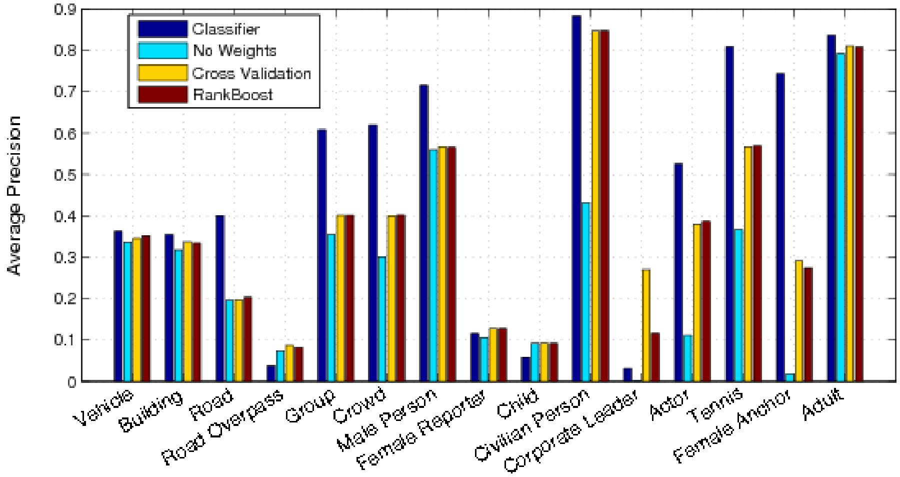

Fig. 2. Concept scalability experiment. Bars display the Average Precision attained by the Columbia classifier, our definition without using fuzzy weights, (i.e., with f_i's set to 1), and our definition using weights calculated using the two proposed methods respectively.

The goal here is to enhance the retrieval performance of the classifiers, using a kind of knowledge-based query expansion.

The average precision attained in this case for the concepts of figure 2 is illustrated in figure 3.

As it can be seen, the results of our method provide improvement over the ones given by the Columbia classifiers. Once again, proper fuzzy weights seem to increase the performance. Finally, comparing these to the results of section 4.1 confirms our expectation that the use of classifiers, whenever available, is beneficial for our method.

4.3 Comparison of Fuzzy Norms

Finally, the same experiment was carried for multiple fuzzy T-norms coupled by their corresponding, dual in terms of fuzzy complement, T-conorms (see [15] for more on this subject). Table 1 displays the mean average precision in each case.

Some comments are worth to be made here: The pair algebraic product/ sum yields the best results in this dataset, while drastic product/ sum seem to be a completely inadequate choice. The standard (min/max) operators have decent, but far from optimal, performance.

Furthermore, contrary to one might expect, the parametric norms (Dubois-Prade and Yager class) have not performed very well. A potential reason might be that in this case optimization may have failed to train their extra parameter over the dataset.

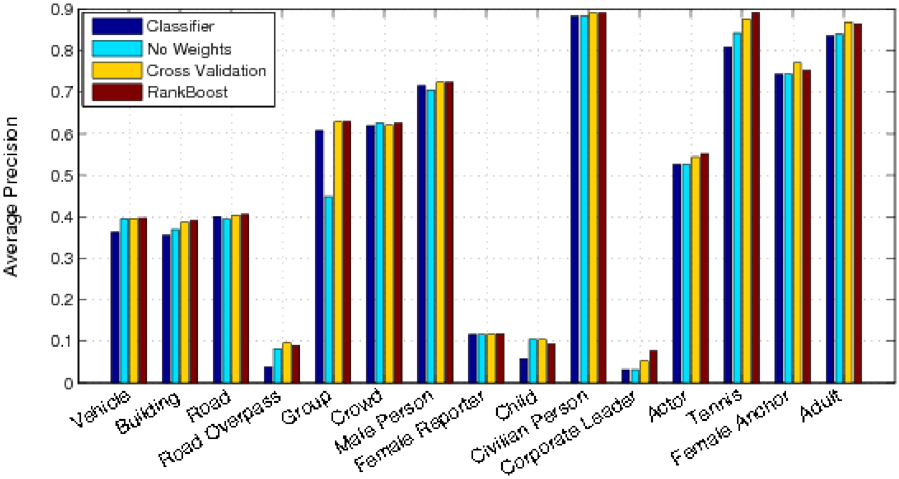

Fig. 3. Classifier enhancement experiment. Bars display the Average Precision attained by the Columbia classifier, our definition without using fuzzy weights, (i.e., with f_i's set to 1), and our definition using weights calculated using the two proposed methods respectively.

Table 1. Mean Average Precision attained for various pairs of fuzzy norms

fuzzy T-norm	cross validation	RankBoost
Standard (min)	0.3065	0.3013
Algebraic product	0.3220	0.3206
Drastic product	0.0866	0.0871
Bounded difference	0.3132	0.3093
Dubois-Prade	0.3087	0.2889
Yager	0.2756	0.2970

5 Conclusions

We have presented a methodology for constructing fuzzy definitions and employing them for concept detection in multimedia, based on classifier results. The fuzzy weights are efficiently adapted to the corpus available. Our approach is useful both extending a concept collection and improving classifier scores in multimedia retrieval, all with a minimal computational effort. Experiments have shown that the method performs well on real data, often outperforming specifically trained discriminative classifiers.

Further improvement of the definition extraction methodology, fine tuning the optimization procedures as well as experimentation on other datasets are of potential interest for the future.

Acknowledgement

This paper is part of the 03ED853 research project, implemented within the framework of the "Reinforcement Programme of Human Research Manpower" (PENED) and co-financed by National and Community Funds (25% from the Greek Ministry of Development-General Secretariat of Research and Technology and 75% from E.U.-European Social Funding).

References

1. Smeaton, A.F., Over, P., Kraaij, W.: Evaluation campaigns and trecvid. In: MIR 2006: Proceedings of the 8th ACM international workshop on Multimedia information retrieval, pp. 321–330. ACM, New York (2006)
2. Snoek, C.G.M., Worring, M., van Gemert, J.C., Geusebroek, J.-M., Smeulders, A.W.M.: The challenge problem for automated detection of 101 semantic concepts in multimedia. In: MULTIMEDIA 2006: Proceedings of the 14th annual ACM international conference on Multimedia, pp. 421–430. ACM, New York (2006)
3. Yanagawa, A., Chang, S.-F., Kennedy, L., Hsu, W.: Columbia university's baseline detectors for 374 lscom semantic visual concepts. Technical report, Columbia University ADVENT Technical Report #222-2006-8 (March 2007)
4. Hauptmann, A., Yan, R., Lin, W.-H., Christel, M., Wactlar, H.: Filling the semantic gap in video retrieval: An exploration. Semantic Multimedia and Ontologies, 253–278 (2008)
5. Christel, M., Hauptmann, A.: The use and utility of high-level semantic features in video retrieval. Image and Video Retrieval, 134–144 (2005)
6. Volkmer, T., Natsev, A.: Exploring automatic query refinement for text-based video retrieval, July 2006, pp. 765–768 (2006)
7. Neo, S.-Y., Zhao, J., Kan, M.-Y., Chua, T.-S.: Video retrieval using high level features: Exploiting query matching and confidence-based weighting, pp. 143–152 (2006)
8. Baader, F., Calvanese, D., McGuinness, D.L., Nardi, D., Patel-Schneider, P.F. (eds.): The Description Logic Handbook: Theory, Implementation, and Applications. Cambridge University Press, Cambridge (2003)
9. Straccia, U.: Reasoning within fuzzy description logics. Journal of Artificial Intelligence Research (April 14, 2001)
10. Stoilos, G., Stamou, G., Tzouvaras, V., Pan, J.Z., Horrocks, I.: A fuzzy description logic for multimedia knowledge representation. In: Proc. of the International Workshop on Multimedia and the Semantic Web (2005)
11. Athanasiadis, T., Simou, N., Papadopoulos, G., Benmokhtar, R., Chandramouli, K., Tzouvaras, V., Mezaris, V., Phiniketos, M., Avrithis, Y., Kompatsiaris, Y., Huet, B., Izquierdo, E.: Integrating image segmentation and classification for fuzzy knowledge-based multimedia indexing. In: Huet, B., Smeaton, A., Mayer-Patel, K., Avrithis, Y. (eds.) MMM 2009. LNCS, vol. 5371, pp. 263–274. Springer, Heidelberg (2009)
12. Mosteller, F.: A k-sample slippage test for an extreme population. The Annals of Mathematical Statistics 19(1), 58–65 (1948)

13. Freund, Y., Iyer, R., Schapire, R.E., Singer, Y.: An efficient boosting algorithm for combining preferences. J. Mach. Learn. Res. 4, 933–969 (2003)
14. Naphade, M., Smith, J.R., Tesic, J., Chang, S.-F., Hsu, W., Kennedy, L., Hauptmann, A., Curtis, J.: Large-scale concept ontology for multimedia. IEEE MultiMedia 13(3), 86–91 (2006)
15. Klir, G.J., Yuan, B.: Fuzzy Sets and Fuzzy Logic; Theory and Applications. Prentice-Hall, Englewood Cliffs (1995)