Minimizing Uncertainty in Semantic Identification When Computing Resources Are Limited

Manolis Falelakis[1], Christos Diou[1],
Manolis Wallace[2], and Anastasios Delopoulos[1]

[1] Department of Electrical and Computer Engineering,
Aristotle University of Thessaloniki - Greece
{fmanf, diou}@olympus.ee.auth.gr,adelo@eng.auth.gr
[2] Department of Computer Science,
University of Indianapolis, Athens Campus - Greece
wallace@uindy.gr

Abstract. In this paper we examine the problem of automatic semantic identification of entities in multimedia documents from a computing point of view. Specifically, we identify as main points to consider the storage of the required knowledge and the computational complexity of the handling of the knowledge as well as of the actual identification process. In order to tackle the above we utilize (i) a sparse representation model for storage, (ii) a novel transitive closure algorithm for handling and (iii) a novel approach to identification that allows for the specification of computational boundaries.

1 Introduction

During the last years the scientific community has realized that semantic analysis and interpretation not only requires explicit knowledge, but also cannot be achieved solely through raw media processing. For this purpose, multimedia research has now shifted from the query by example approach, where the aim was to provide meaningful handling and access services directly from the low level processing of media, to a two step approach including (i) the identification of high level entities in raw media and (ii) the utilization of this *semantic indexing* towards the offering of more efficient multimedia services. Research efforts in standardizing the metadata representations of multimedia documents have led to the MPEG-7 standard which, however, does not suggest methods for automatic extraction of high level information.

In well-structured specific domains (e.g., sports and news broadcasting), domain-specific features that facilitate the modelling of higher level semantics can be extracted (see e.g., [6]). Typically, a priori knowledge representation models are used as a knowledge base that assists semantic-based classification and clustering [9]. In [7], for example, the task of bridging the gap between low-level representation and high-level semantics is formulated as a probabilistic pattern recognition problem.

W. Duch et al. (Eds.): ICANN 2005, LNCS 3697, pp. 817–822, 2005.

In the proposed paper we aim to deal with the identification of semantic entities in raw media, focusing on issues related to efficient operation under computing resources limitations. The aim is to automatically configure the identification process so as to achieve optimal results, i.e. maximize the relevance of the retrieved multimedia documents when the amount of available computing resources is constrained. Constraints are directly related to hard bounds regarding physical memory, processing power and time availability.

These considerations turn to be important due to two major reasons. The first is that the knowledge base involved in semantic identification procedures naturally contains a vast amount of items so that even simple operations on its content may lead to overwhelming the existing computational power and/or memory. The second reason is that although retrieval is performed on the basis of semantic entities, identification of the latter necessarily resorts to quantification of large numbers of measurable features (syntactic entities) i.e., requires execution of multiple signal and image processing algorithms. The latter may sum up to an execution cost that is prohibitive especially for real-time or online scenarios.

2 Knowledge Representation

In order to extract high level (semantic) information from multimedia, low level (directly measurable) features have to be evaluated and combined with the use of a properly structured knowledge base [10]. This structure is often called "semantic encyclopedia" and consists of relationships either among semantic entities or between semantic entities and low level features [2]. For example semantic entity "planet" can be related with semantic entity "star" while feature "blue color" can be related with semantic entity "sea". In most cases such relationships are valid *up to a certain degree*, that is, there is an inherent uncertainty and/or degree of validity associated with them. This makes representation of the aforementioned relationships by using fuzzy relations a natural choice. Considering the set of both semantic and syntactic entities as our universe of discourse, the semantic encyclopedia can be modelled as a large fuzzy relation describing the degrees of association among the elements of this universe. Using such an encyclopedia, it is possible to build systems for automatic or semi-automatic identification of semantic entities in raw media, thus contributing to the bridging of the semantic gap [5].

Even in the case of semantic encyclopedias that are limited to specific thematic categories (e.g., sports, politics, etc) the number of included "terms" and syntactic features may easily reach the order of tens of thousands (in [2], for example, the universe of discourse contains definitions for 70000 semantic entities). This alone is prohibitive for the complete representation of the semantic relations. On the other hand, classical linked list sparse array representations, as well as hash table approaches, are both inadequate to handle such sizes of data, the former due to the $O(n)$ access time and the latter due to the augmented requirements in physical memory. In this work we utilize a novel sparse fuzzy binary relation model that is based on pairs of AVL trees and provides for both space efficient storage and time efficient access.

Specifically, the representation model proposed in order to overcome these limitations is as follows: a binary relation is represented using two *AVL trees*; an AVL tree is a binary, balanced and ordered tree that allows for access, insertion and deletion of a node in $O(\log m)$ time, where m is the count of nodes in the tree [1]. If $n \log n$ nodes exist in the tree, as will be the case for the typical sparse relation, then the access, insertion and deletion complexity is again $O(\log n)$ since $n < n \log n < n^2 \Rightarrow O(\log n) \leq O(\log(n \log n)) \leq O(\log n^2) = O(\log n)$.

In both trees, both row index i and column index j are utilized to sort the nodes lexicographically; however, the first tree, the row-tree, is sorted according to index i, and in case of common row positions i, column position j is utilized, and vice versa for the second tree, the column-tree. The resulting vectors can then be represented as AVL trees.

Furthermore, most of the semantic relations that participate in the semantic encyclopedia are of a transitive form. For example, a texture feature may be associated with entity "skin" and "skin" with semantic entity "human", which should imply that the specific feature is also associated with entity "human". In order to populate the semantic encyclopedia with the links that can be inferred in this way and to allow for more efficient time wise access to such implied links, a transitive closure of these relations needs to be acquired. As conventional transitive closure algorithms either have a high complexity ($O(n^3)$ or higher) or cannot handle archimedean t-norms and asymmetrical relations, such as the ones typically included in a semantic encyclopedia, this task is not trivial.

Based on the proposed representation model, a novel transitive closure algorithm that is targeted especially to generalized, sparse fuzzy binary relations and has an almost linear complexity can be utilized [8]. This algorithm has the added advantage of allowing updates to the relation with trivial computational burden while at the same time maintaining the property of transitivity.

3 Semantic Identification

The methodologies referred to up to this point guarantee that the encyclopedia will be well structured, consistent and informative enough (due to its transitive closure characteristic) and represented in a compact manner. The remaining of the paper refers to its effective use in identifying semantic entities and retrieving the corresponding multimedia documents. The core idea is that the analysis of the raw media by applying signal and image processing algorithms yields quantification of existence of low level features as a first stage of the retrieval procedure. The second step is the assessment of the degree up to which certain semantic entities are identified within the given multimedia documents by exploiting the relations of the semantic encyclopedia. We choose to model this identification procedure as a fuzzy inference mechanism. Considering, though, the count of semantic entities in the universe of discourse, as well as the count of distinct features that may be extracted and evaluated, it is easy to see that this process quickly becomes inapplicable in real life scenarios.

The way followed in this paper in order to tackle semantic identification, without suffering the expense of immense processing power, is to partially identify an entity i.e., to evaluate only a subset of the involved syntactic features and produce an estimation based on this imperfect and incomplete input. The challenge is to automatically select this subset that includes characteristics providing the highest possible validity regarding the computed result, combined with minimum complexity requirements. Utilization of fuzzy logic theory is shown to provide methods for quantifying both validity and complexity of the involved algorithms.

Evaluation of a Syntactic Entity Y_i participating in a detailed definition is equivalent to running its corresponding algorithm τ and computing the membership degree μ_{Y_i} up to which the document under examination assumes property Y_i. In a similar manner, a metric is defined that denotes the degree up to which a Semantic Entity exists in a document and is called *Certainty* of the identification. Given a detailed definition of a Semantic Entity E_k in the form

$$E_k = F_{1k}/S_1 + F_{2k}/S_2 + \ldots + F_{nk}/S_n,$$

and the membership degrees μ_{Y_i} of the Syntactic Entities Y_i in a specific document, Certainty that E_k exists in that document is defined as

$$\mu_{E_k} \triangleq \mathcal{U}_i(\mathcal{I}(F_{Y_i E_k}, \mu_{Y_i}))$$

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

The maximum possible value of μ_{E_k} is assigned the term *Validity* of the definition and is equal to

$$\mathcal{V}(E_k) \triangleq \mathcal{U}_i(F_{Y_i E_k}),$$

attained for $\mu_{Y_i} = 1$ for all Y_i in the scope of E_k and the use of the identity $I(a, 1) = a$ (true for every t-norm $\mathcal{I}$).

Validity denotes the maximum amount of information that a definition can provide and is used extensively in the identification design process. We must note that Validity is independent of the data set under examination and can be computed *prior* to the identification. Validity is therefore a property of the definition itself.

Another characteristic of a definition is the computational complexity associated with the algorithms corresponding to its Syntactic Entities. We assign a computational cost $c(t)$ to every syntactic feature t that is essentially equal to the cost of it's corresponding algorithm τ. Hence, we may now define *Complexity* of a definition as

$$\mathcal{C}(E_k) = \sum_i c(t_i)$$

where t_i are the syntactic features required to evaluate the properties Y_i of the definition E_k. Notice that this value will normally depend on the size of the input data, as will the values $c(t_i)$. At least, though, worst or average case expressions of $c(t_i)$ can be considered as independent of the actual content of the examined data sets. In this perspective $\mathcal{C}(E_k)$ is also computable prior to identification.

Based on these definitions, and following the approach of [3] and [4], the task of optimal semantic identification given some hard complexity boundaries is reduced to a dynamic programming optimization problem. Thus, semantic identification can be performed in real-time, while the uncertainty in the output is guaranteed to be minimized.

4 Experimental Results

In this section we provide some brief, yet indicative, results acquired through the application of the proposed methodologies.

Representation. A knowledge of 70000 semantic entities has been developed [2]. Although loading a fuzzy binary relation defined on this set requires more than 50GB of main memory, assuming a double precision number format, the knowledge base is loaded in less that 100MB of memory using the proposed representation model.

Handling. Transitive closure of the above-mentioned relation is calculated to require more than 5 days of computing time. Using the proposed representation model and applying the proposed transitive closure algorithm, the time required for the complete transitive closure is approximately 20 seconds on the same computer. For the case of simple update of an already transitive relation, the processing time is less than a millisecond.

Identification. Figure 1 presents the validity achieved when applying the prosed approach on a random data set with different complexity thresholds. The non linear character of the graph shows the benefit of the optimized selection of the part of the definition to evaluate in each case. Note that for a threshold $C_T = 12$ the Validity is $V \approx 0.9$ while the total Complexity of the definition is $C_t = 132$ for the uniformly distributed values.

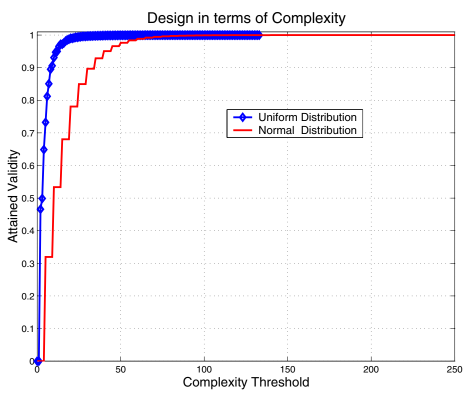

Fig. 1. Achieved validity with respect to the complexity threshold

5 Conclusions

In this paper we made a first attempt to provide an integrated solution to computing problems related to semantic identification of entities in multimedia streams. Problems addresses include the storage requirements of the semantic knowledge, the computational complexity for the handling of the knowledge, as well as the computational complexity of the identification process itself.

The above have been tackled through a novel representation model and matching algorithms for transitive closure and semantic identification. The experimental results verify the efficiency and prospect of the proposed approaches.

References

1. Adelson-Velskii G.M., Landis E.M., "An algorithm for the organization of information" Doklady Akademia Nauk SSSR, Vol 146, pp. 263-266, 1962; English translation in Soviet Math, Vol 3, pp. 1259-1263, 1962.
2. Y. Avrithis, G. Stamou, M. Wallace, F. Marques, P. Salembier, X. Giro, W. Haas, H. Vallant and M. Zufferey, "Unified Access to Heterogeneous Audiovisual Archives", Journal of Universal Computer Science Vol 9(6), pp. 510-519, 2003.
3. M. Falelakis and C. Diou and A. Valsamidis and A. Delopoulos, "Complexity Control in Semantic Identification", IEEE International Conference on Fuzzy Systems, Reno, Nevada, USA, May 2005.
4. M. Falelakis and C. Diou and A. Valsamidis and A. Delopoulos, "Dynamic Semantic Identification with Complexity Constraints as a Knapsack Problem", IEEE International Conference on Fuzzy Systems, Reno, Nevada, USA, May 2005.
5. X. Giro and F. Marques, "Semantic Entity Detection using Description Graphs", International Workshop on Image Analysis for Multimedia Interactive Services (WIAMIS), London, England, April 2003.
6. W. Al-Khatib and Y.F. Day and A. Ghafoor and P.B. Berra, "Semantic Modeling and Knowledge Representation in Multimedia Databases", IEEE Transactions on Knowledge and Data Engineering, Vol 11(1), pp. 64-80, 1999.
7. M. Ramesh Naphade and I.V. Kozintsev and T.S. Huang, "A factor graph Framework for Semantic Video Indexing", IEEE Trans. on Circuits and Systems for Video Technology, Vol 12(1), pp. 40-52, 2002.
8. M. Wallace, Y.Avrithis and S. Kollias, " Computationally efficient sup-t transitive closure for sparse fuzzy binary relations", Fuzzy Sets and Systems, in press.
9. A. Yoshitaka and S. Kishida and M. Hirakawa and T. Ichikawa, "Knowledge-assisted content based retrieval for multimedia databases" IEEE Multimedia, Vol. 1(4), pp. 12-21, 1994.
10. R. Zhao and W.I. Grosky, "Narrowing the Semantic Gap-Improved Text-Based Web Document Retrieval Using Visual Features", IEEE Transactions on Multimedia, Vol 4(2), 2002.