A Framework for Image Annotation Using Semantic Web

Ahmed Bashir and Latifur Khan

University of Texas at Dallas
{ahmedb, lkhan}@utdallas.edu

Abstract. The impetus behind Semantic Web research remains the vision of supplementing availability with utility; that is, the World Wide Web provides availability of digital media, but the Semantic Web will allow presently available digital media to be used in unseen ways. An example of such an application is multimedia retrieval. At present, there are vast amounts of digital media available on the web. Once this media gets associated with machine-understandable metadata, the web can serve as a potentially unlimited supplier for multimedia web services, which could populate themselves by searching for keywords and subsequently retrieving images or articles, which is precisely the type of system that is proposed in this paper. Such a system requires solid interoperability, a central ontology, semantic agent search capabilities, and standards. Specifically, this paper explores this cross-section of image annotation and Semantic Web services, models the web service components that constitute such a system, discusses the sequential, cooperative execution of these semantic web services, and introduces intelligent storage of image semantics as part of a semantic link space.

1 Introduction

The impetus behind Semantic Web research remains the vision of supplementing availability with utility; that is, the World Wide Web provides availability of digital media, but the Semantic Web will allow presently available digital media to be used to serve new purposes, an example of which is image retrieval.

The Semantic Web is an extension of today's Web technology; it boasts the ability to make Web resources accessible by their semantic contents rather than merely by keywords and their syntactic forms. Due to its well-established mechanisms for expressing machine-interpretable information, information and Web services previously available for human consumption can be created in a well-defined, structured format from which machines can comprehend, process and interoperate in an open, distributed computing environment.

This proves to be quite advantageous with respect to data collection; intelligent software entities, or agents, can effectively search the web for items of interest, which they can determine with new semantic knowledge. For instance, sports images or articles can be retrieved from around the web and processed by the respective web services to enhance a website in terms of the sheer multimedia content available. In such a system, the semantic web serves as a large, automated image collection that may be used to populate an annotated image "gallery". This image "gallery" would be represented as a semantic link space that organizes like images together based on known image semantics; for example, all basketball images would be grouped as an image network, and so on.

Combining image retrieval with the Semantic Web, however, is not merely beneficial due to the availability of raw data or the potential for automated image annotation, but there is also the added benefit of using a web ontology, or a set of concepts and their interrelations. By using such ontologies not only to search for multimedia but also to classify it, the system ensures consistency in terminology, leading to more accurate and precise query results [1, 7, 8, 9, 10].

1.1 The Approach

At present, there are vast amounts of digital media available on the web. Once this media gets associated with machine-understandable metadata, the web can serve as a potentially unlimited supplier for multimedia web services, which could populate themselves by searching via keywords and subsequently retrieving images or articles. This presents a novel approach to semi-automatic image annotation or classification. In this case, not only is the annotation done automatically once both the support vector machine and the Bayesian network are trained, but the source is replenished automatically, as well.

The image annotation task has been decomposed into classification of low-level, or atomic, concepts and classification of high-level concepts in a domain-specific ontology. In the general sense, concepts are atomic if they are terms that can describe specific objects or image segments. Examples would be ball, stick, net, and other well-defined objects. High-level semantic concepts, on the other hand, are used to describe an environment with a set of existing atomic concepts associated to it. For example, an image that contains a ball, a net, shoes, and humans can be described as a basketball game. The framework takes advantage of this natural gap in semantics, classifying atomic concepts using support vector machines and high-level concepts using Bayesian belief networks.

Upon classifying the image, the system would reflect the image semantics, its features, content, and semantic category, as part of a semantic space. Figure 1 illustrates the layered architecture. The bottom-most layer represents the original image, and the layer directly above it will represent the image semantics using an ontology. The semantic space can then, as mentioned, prepare image networks based on the available image semantics, and the features that correspond to the respective images will constitute the feature space. As part of the operation interface, a user or a web agent can query the system, which would search and retrieve image information from the underlying layer [9].

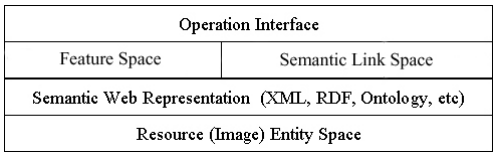

Fig. 1. Semantic space architecture

1.2 Experimental Context

Results have shown that separating atomic classification from high-level classification improves Bayesian classification by reducing the complexity of the directed acyclic graph associated with the Bayesian network. This way, image features are abstracted away and the Bayesian structure includes only semantic concepts. With a reasonably strong segmentation algorithm, results are promising. To test the strengths and weaknesses of this system, a classic segmentation algorithm has been combined with a support vector machine and a Bayesian network. The training set consists of 3,000 feature sets, and over 300 images have been classified.

1.3 Contributions

This paper explores this cross-section of image annotation and Semantic Web services, models the web service components that constitute such a system, discusses the sequential, cooperative execution of these semantic web services, and presents the technical challenges. The main contributions deal with the integration of these new service-building technologies, the use of two classification methods to separate high-level and low-level semantic concepts, the hyperlink-based search and collection of fresh raw images using intelligent web agents, and of course the representation of key image information in terms of a semantic link space rather than a local image repository. Another key contribution is the use of a web ontology as a multipurpose tool that seamlessly integrates different service components; the ontology can serve as the Bayesian structure to classify images or text, a translator to understand user queries, and an instructor for agents that gather multimedia.

2 Proposed Architecture for Prototype System

2.1 Framework Description

The prototype system will consist of an interface through which users can query the system regarding specific sports. This request would be processed by a service that would retrieve the relevant images and articles from their respective repositories and present them to the user.

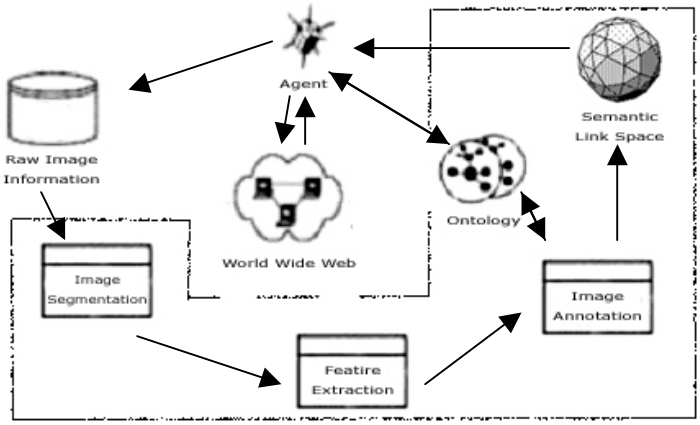

Fig. 2. Proposed architecture

As depicted in Figure 2, web agents collect unclassified images using a hyperlink-based approach in order to build the semantic link space containing image semantics and classifications; agents will discard images that cannot be supported by the system. Supported images are segmented into various objects, and the objects' features are subsequently extracted; features include hue, intensity, saturation, and shape. Feature vectors are sent into the support vector machine, which will identify low-level concepts that exist in each image. The set of low-level concepts are sent as known evidence into a trained Bayesian network, which will classify the images according to high-level semantic concepts. The image semantics can then be stores as part of the semantic link space.

Lastly, the system contains the central ontology. This ontology will contain detailed information regarding the sports domain, specifically the different types of sports, the equipment involved, and so on. Moreover, the service ontology will be continually changing as agents are able to discover more sports and so on. For example, agents may discover a new sport, tennis for example, and add it into the ontology. Alternatively, agents may find additional gear that is associated to an existing sport, such as a baseball helmet. However, the system as discussed in this paper assumes a single ontology without any ontology merging, which is beyond the scope of this paper.

2.2 Technical Challenges

To implement such a system involves an integration of several up and coming technologies. This section highlights some of the technologies involved and the challenges presented by each of them.

2.2.1 Image Annotation

Annotation has been decomposed into identifying low-level and high-level semantic concepts, respectively. The former will be determined using support vector machines, and Bayesians networks will determine the latter.

The support vector method, or SVM, is a technique which is designed for efficient multidimensional function approximation; the aim is to determine a classifier or regression machine which minimizes the training set error [6]. The basic procedure is to fix the empirical risk associated with an architecture and then to use a method to minimize the generalization error. The primary advantage of support vector machines as adaptive models for binary classification and regression is that they provide a classifier with a low expected probability of generalization errors. This approach can be trivially extended to multi-class classification by getting a binary response with respect to each atomic classification.

Bayesian networks are used to model the causal relationships that exist in the context of image semantics [3]. This will allow a system to associate query keywords with other semantic concepts to some degree of belief. Hence, image queries will be more intelligently handled and will yield better results, a direct result of understanding the dependencies and relationships between different semantic concepts. The causal relationships can be patterned using the provided web ontology, which already represents a set of concepts and their interrelations. The variables that will make up the network will be a combination of high-level semantic

classifiers as well as atomic level classifiers sent from the Support Vector Machine. For the purpose of this project, the Bayesian structure is precisely the web ontology.

2.2.2 OWL-S

To make use of a Web service, a software agent needs a computer-interpretable description of the service, and the means by which it is accessed. Semantic Web markup languages must not only establish a framework within which these descriptions are made and shared but also enable one web service to automatically locate and utilize services offered by other web services [5]. OWL-S provides the solution, providing facilities for describing service capabilities, properties, pre-/post-conditions, and input/output specifications.

2.2.3 WSDL

Web Services Description Language (WSDL) is a new specification to describe networked XML-based services. It provides a simple way for service providers to describe the basic format of requests to their systems regardless of the underlying protocol, in our case SOAP. Under the WSDL standard, network services are viewed as a set of endpoints operating on messages containing either document-oriented or procedure-oriented information. The operations and messages are described abstractly, and then bound to a concrete network protocol and message format to define an endpoint. Related concrete endpoints are combined into abstract endpoints, or services. WSDL is extensible to allow description of endpoints and their messages regardless of what message formats or network protocols are used to communicate [4].

WSDL documents describe operations, messages, datatypes, and communication protocols specific to a web service. To carry out the communication between web services, SOAP will be used. SOAP provides the framework by which application-specific information may be conveyed in an extensible manner. Also, SOAP provides a full description of the required actions taken by a SOAP node on receiving a SOAP message. The SOAP stack will convert SOAP requests into native requests that the web service can make use of. Similarly, the web services' responses must be designed as SOAP responses.

2.2.4 Semantic Link Space

Information regarding classified images will be organized using a semantic link space [10]. By associating like images together, image networks are created (see Figure 1); similarity will be judged based on image semantics, including hue, saturation, intensity, and shape. By using this idea in conjunction with the hyperlink-based approach, user queries would be satisfied.

3 Results and Conclusions

Results from image annotation have proved that a properly trained support vector machine and Bayesian network can work alongside one another to produce satisfactory results. The SVM was trained with a mix of basketball, baseball, bat, soccer, hoop, and grass images that total 3000 training objects. The training set captured key characteristics of each image segment: hue, saturation, intensity, and shape. Once the support vector machine was trained, the training set was also used as test data in order to judge the training accuracy, which averaged at 98.5%.

The recall values indicate how well the system fared in recognizing all the segments that depict the same object. For instance, in the case of grass, the system is able to recognize 74 of the 80 grass objects, resulting in a 93% recall. However, there are 88 grass objects recognized, so the precision value is used to indicate how many of the retrieved positively classified images truly depict the object in question. In this case, 74 of the 88 positively classified grass objects are actually grass objects, leading to an 84% precision. Some of the problems with atomic classification are intuitive. In the case of a basketball hoop, the support vector machine is attempting to recognize an object that lacks a definite shape or color. Soccer balls are also tough to recognize because they are composed of two distinct colors: black and white. Complete results are presented in Tables 1 and 2.

Table 1. SVM classification results

Object	# Expected	# Retrieved	Recall	Precision
Soccer	80	61	39%	51%
Grass	80	88	93%	84%
Basketball	150	134	89%	100%
Baseball	100	123	83%	67%
Bat	80	81	100%	99%
Hoop	30	81	100%	37%

Table 2. Bayesian classification results

Classification	# Expected	# Retrieved	Recall	Precision
Basketball	150	181	92.7%	76.8%
Soccer game	80	111	100%	72.1%
Baseball game	100	135	96%	71.1%

Higher-level classification suffers in all instances where atomic classification falls short; however, one idea that deserves mention is the difference between misclassification and unclassification. For example, even if a basketball is not recognized as a basketball, there is still an inherent benefit of not recognizing the basketball as another type of object, a soccer ball for instance. In the case of the 150 basketball pictures, 136 were retrieved due to limited misclassification. On the other hand, of the 80 soccer images, 35 were misclassified at the atomic level as containing baseballs. In this case, the Bayesian network will be unable to accurately classify the image, which will subsequently be discarded.

The results, particularly the precision values, show that there are too many multiple classifications. For example, an image that contains a soccer ball, a bat, and a baseball will be retrieved both as a baseball image as well as a soccer image. Another important note is that, due to a simple Bayesian structure, images were often classified correctly if one of two objects were recognized.

The system will be extended to recognize details pertaining to the environment so images can be classified on the basis of indoors or outdoors and team or individual. Extracting such detailed information from an image will again require a strong segmentation algorithm coupled with some preprocessing that will give way to more intelligent segmentation so that regions can be identified more accurately.

References

1. O. Marques and N. Barman. Semi-automatic Semantic Annotation of Images Using Machine Learning Techniques. In *The Semantic Web - ISWC 2003 Proceedings*, pages 550-565, 2003.
2. Protégé-2000. http://protege.stanford.edu/
3. R. E. Neapolitan. *Learning Bayesian Networks*. Prentice-Hall, Upper Saddle River, NJ, 2004.
4. E. Christensen, F. Curbera, G. Meredith, and S. Weerawarana. Web Services Description Language (WSDL) 1.1. http://www.w3.org/TR/wsdl
5. The OWL Services Coalition. OWL-S: Semantic Markup for Web Services. http:// www.daml.org/ services/owl-s/1.0/owl-s.html
6. C. Cortes and V. Vapnik. Support vector networks. *Machine Learning*, 20:1-25, 1995.
7. L. Khan, D. McLeod, and E. Hovy. Retrieval Effectiveness of Ontology-based Model for Information Selection. The VLDB Journal: The International Journal on Very Large Databases, ACM/Springer-Verlag Publishing, Vol. 13(1): 71-85 (2004).
8. A. Benitez and Shih-Fu Chang. Multimedia Knowledge Integration, Summarization and Evaluation. http://citeseer.ist.psu.edu/benitez02multimedia.html"
9. Hai Zhuge. Semantic-Based Web Image Retrieval. http://www2003.org/cdrom/papers/poster/p172/p172-zhuge/p172-zhuge.htm
10. Eero Hyvönen, Samppa Saarela, Avril Styrman, and Kim Viljanen. Ontology-Based Image Retrieval. http://www.cs.helsinki.fi/group/seco/presentations/www2003/p199-hyvonen.html