Supervised Information Retrieval for Text and Images

Alexandros Kyriakides, Giorgos Zacharia

The Problem: To perform retrieval of images. We will use the captions of the images that appear in news articles. Given a news story that appears in a news article, our goal is to select an appropriate image for that article. The problem is to select the most important features (or words) from the news story that will help us select the correct image.

Motivation: In recent years, with the continuing increase in available storage space, image repositories have grown to such sizes that it is becoming ever more difficult to keep track of this large amount of data. These images are usually "unlabeled." If these images were categorized and "labeled," it would allow for the simple retrieval of a specific image upon request. It is important therefore, to find methods which enable us to automatically categorize and label these images. Say for example, we were requesting an image (photo) of the president of the U.S.A. from a large database of images. If all the images had already been placed into categories and labeled, then it would be trivial to just return all the images which fall into the category of U.S. President. The motivation for this work is the fact that several learning-machine algorithms have already been developed. Our goal is to combine several such algorithms and to use boosting techniques to improve our results.

Previous Work: As stated in [7], Image Retrieval has been a very active research area since the 1970's. Image Retrieval has been studied from two different angles: text-based and visual-based. The fields of text mining, boosting and combination of learning machines are of particular interest to us. Text mining describes the technique of extracting useful data from relatively unstructured collections of text. Boosting is a general method which attempts to "boost" the accuracy of any given learning algorithm. (Schapire [8]) . Schapire[9] came up with the first provable polynomial-time boosting algorit hm in 1989. Algorithms have also been developed that make decisions by combining the predictions of several prediction strategies, called experts. [1] Some algorithms also use the added feature of allo wing experts to abstain from making a prediction. Such experts are called specialists, because we think of them as making their prediction only when the instance to be predicted falls within their area of expertise [4]. Multi-label text classification methods have been studied and analyzed by McCallum [6] and by Imai et al. [5]. Huang et al. [4] state that not much work has been done on automatic im age classification, and propose a method for hierarchical classification of images via supervised learning.

Approach: Currently our work focuses on the news story text and the captions of images. We have accumulated a database of 25 000 news articles that have appeared on online news sources. Each news article is accompanied with one or two images. Each image has a caption of approximately 50 words. The main difficulty we face is the large number of classes which will exist in this multi-class classification. Even by clustering the images based on text similarity, the number of distinct classes will still be large. To avoid a large number of classes, we performed a binary classification for each word appearing in a news story. A word falls into the "yes" class if it appears in both the news story and the accompanying image caption. Otherwise, it falls into the "no" class. A feature vector was calculated for each word. We currently used 10 features for each word, including "part of speech", "TFIDF score", "Infogain", and position of word in story. A description of TFIDF and Infogain can be found in [2]. The results were encouraging. Using boosting algorithms, we managed to create a model that on the average gave approximately 80% accuracy on predicting if a word in a news story should be classified as "yes" or "no". This model can now be used to extract important words from news stories in order to make the original classification task more feasible. The confusion matrix below shows some of the results.

Difficulty: The main difficulty is the large number of classes that exist. In order to make the problem less difficult we are currently performing binary classification on the words of the news story. This has proven very effective.

Impact: The direct impact is that it will be possible to select images for news articles with minimum human intervention. Also, with the approach we are taking, we will be able to select important features (words) from news stories. These features will can also be used for summarization purposes and for document retrieval systems.

Research Support: This report describes research done at the Center for Biological & Computational Learning, which is in the Department of Brain & Cognitive Sciences at MIT and which is affiliated with the McGovern Institute of Brain Research and with the Artificial Intelligence Laboratory.

This research was sponsored by grants from: Office of Naval Research (DARPA) Contract No. N00014-00-1-0907, Office of Naval Research (DARPA) Contract No. N00014-02-1-0915, National Science Foundation (ITR/IM) Contract No. IIS-0085836, National Science Foundation (ITR/SYS) Contract No. IIS-0112991, National Science Foundation (ITR) Contract No. IIS-0209289, National Science Foundation-NIH (CRCNS) Contract No. EIA-0218693, and National Science Foundation-NIH (CRCNS) Contract No. EIA-0218506.

Additional support was provided by: AT&T, Central Research Institute of Electric Power Industry, Center for e-Business (MIT), DaimlerChrysler AG, Compaq/Digital Equipment Corporation, Eastman Kodak Company, Honda R&D Co., Ltd., ITRI, Komatsu Ltd., The Eugene McDermott Foundation, Merrill-Lynch, Mitsubishi Corporation, NEC Fund, Nippon Telegraph & Telephone, Oxygen, Siemens Corporate Research, Inc., Sony MOU, Sumitomo Metal Industries, Toyota Motor Corporation, and WatchVision Co., Ltd.

Results:

Confusion Matrix:

		Predicted Class	
		No	**Yes**
Actual Class	**No**	120618	32040
	Yes	4771	8764

Correctly Classified Instances: 129382 (77.9 %)

References:

[1] Nicol Cesa-Bianchi, Yoav Freund, David P. Helmbold, David Haussler, Robert E. Schapire, and Manfred K. Warmuth. How to use expert advice. *Journal of the Association for Computing Machinery, 44(3):* 427-485, 1997.

[2] Robert Cooley, "Classification of news stories using support vector machines," in IJCAI'99 Workshop on Text Mining, Stockholm, Sweden, Aug. 2 1999.

[3] Yoav Freund, Robert E. Schapire, Yoram Singer and Manfred K. Warmuth. Using and combining predictors that specialize. In *Proceedings of the Twenty-Ninth Annual ACM Symposium on the Theory of Comput ing, pages 334–343,* 1997.

[4] J. Huang, S. R. Kumar, and R. Zabih. An automatic hierarchical image classification scheme. In Sixth ACM Intl. Multimedia Conf., pages 219–228, 1998.

[5] Toru Imai, Richard Schwartz, Francis Kubala, and Long Nguyen. Improved topic discrimination of broadcast news using a model of multiple simultaneous topics. In *Proceedings of the IEEE ICASSP*, volum e 2, pages 727-730, Munich, Germany, April 1997.

[6] Andrew McCallum. Multi-Label Text Classification with a Mixture Model Trained by EM. Revised version of paper appearing in AAAI'99 *Workshop on Text Learning.*

[7] Yong Rui, Thomas S. Huang and Shih-Fu Chang. Image retrieval: Past present and future. Journal of Visual Communication and Image Representation, Volume 10, pages 1 – 23, 1999.

[8] Robert E. Schapire. A brief introduction to boosting. In *Proceedings of the Sixteenth International Joint Conference on Artificial Intelligence*, 1999.

[9] Robert E. Schapire. The strength of weak learnability. *Machine Learning.* 5(2):197-227;1990.