Mining multimedia salient concepts for incremental information extraction

Conference or Workshop Item

Mining Multimedia Salient Concepts for Incremental Information Extraction

João Magalhães
Imperial College London
Department of Computing
South Kensington Campus
London, UK
j.magalhaes@imperial.ac.uk

Stefan Rüger
Imperial College London
Department of Computing
South Kensington Campus
London, UK
s.rueger@imperial.ac.uk

ABSTRACT
We propose a novel algorithm for extracting information by mining the feature space clusters and then assigning salient concepts to them. Bayesian techniques for extracting concepts from multimedia usually suffer either from lack of data or from too complex concepts to be represented by a single statistical model. An incremental information extraction approach, working at different levels of abstraction, would be able to handle concepts of varying complexities. We present the results of our research on the initial part of an incremental approach, the extraction of the most salient concepts from multimedia information.

Categories and Subject Descriptors
I.2.10 [**Vision and Scene Understanding**]: Video analysis; H.3.1 [**Content Analysis and Indexing**]: Abstracting methods; H.3.3 [**Information Search and Retrieval**]: Clustering.

General Terms
Algorithms, Measurement, Experimentation.

Keywords
Multimedia information extraction, multimedia clustering.

1. INTRODUCTION
The first multimedia information extraction algorithms for retrieval applications have used Bayesian techniques for extracting concepts of different complexities (e.g. *sky vs. pencil*) [1]. Most of the problems in multimedia information extraction reside in the degree of ambiguity of most concepts versus the available training data. Some concepts have an audio-visual representation too complex to be captured by a single model: the confusion between concepts is reflected on the high degree of uncertainty of the extracted information, [2]. Some approaches formulate the information extraction problem as a cross-lingual retrieval problem and proposed a multimedia equivalent solution, the cross-media relevance model [3].

One might suppose that for some concepts the learning task is just too difficult to be accomplished. For example, concepts such as *sun*, *outdoor* or *indoor*, may be easy to detect, but concepts such as *bird*, *plane* or *superman*, may be more reliably detected if other, more basic/salient, concepts were detected previously. In an incremental information extraction approach, the 'trustworthiness' of this salient extracted information is crucial for the success of the algorithms to extract complex concepts. So, we propose a novel algorithm for extracting salient concepts by mining the feature space clusters and then assigning their labels.

SIGIR'05, August 15–19, 2005, Salvador, Brazil.
ACM 1-59593-034-5/05/0008.

2. ALGORITHM
As an alternative to learning the models of the concepts in a high-dimensional feature space, we 'mine' the feature space for salient patterns and then label those patterns (or semantically correlated patterns) with the corresponding salient concept. With this novel strategy we intend to achieve an extraction algorithm that is not too attached to the training examples of a specific concept but to the most basic/salient patterns present in the multimedia dataset.

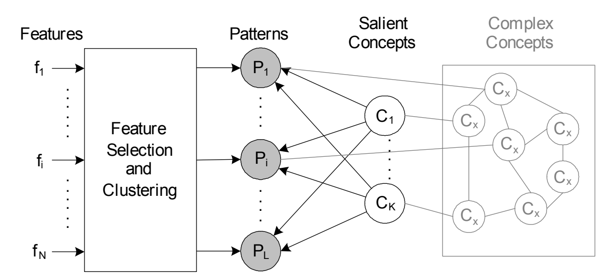


Figure 1. The incremental information extraction framework.

The salient concepts extraction algorithm is divided into the following smaller tasks, Figure 1:

1. Features Pre-Processing: Process the features, by removing non-relevant features, combining redundant ones, and normalizing them.
2. Clustering the Feature Space: We use an unsupervised learning algorithm (finite mixture clustering) to detect salient patterns in the feature space.
3. Learning Pattern-Concept Models: The final step is to model the relation between the detected patterns and concepts with a Bayesian network.

The following sections describe the details of the algorithm.

2.1 Feature Pre-Processing
Little work has been done in feature selection algorithms for unsupervised learning. It is especially hard to compute the weights of each feature if one doesn't know the classes' labels. Possible criteria for feature selection are based on feature similarity or on clusters *separability* measures. The first type

(filter methods) use a feature dimension similarity measure to decide which dimensions are merged or removed (e.g. PCA [4], ICA [4], cross-entropy based algorithm proposed by Koller et al. [5]). The second type (wrapper methods) involves embedding the feature selection with the clustering algorithm.

2.2 Clustering the Feature Space

In the algorithm's second step, the feature space clustering is done under the assumption that features are independent. That is, each feature is processed individually. We use the finite-mixture of Gaussians clustering proposed in [6]. The finite-mixture learning algorithm follows the EM algorithm with embedded model selection using a Minimum Message Length criterion. This algorithm is based on EM and avoids some of its drawbacks: sensitivity to initialization, possible convergence to the boundary of the parameter space, and its the estimation of different feature importance. The algorithm is also capable of selecting the mixture's number of components.

After learning the finite-mixture model we consider each Gaussian component to be a salient pattern (a cluster), $\{P_1, \ldots, P_L\}$.

2.3 Learning Pattern-Concept Models

So far, we have ignored the labels of the training examples: the algorithm worked completely unsupervised and with the entire training set. Now, we use the labels to learn a Bayesian network, [7]: the relations between patterns $\{P_1, \ldots, P_L\}$ and concepts $\{C_1, \ldots, C_K\}$, and the parameters of the concepts' nodes.

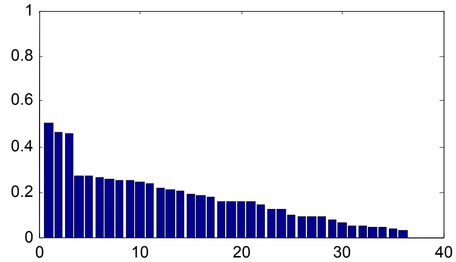
We used the K2 algorithm to search the Directed Acyclic Graph space [7]. The DAG space is composed by DAGs which only have edges between pattern nodes and concepts nodes are considered (we ignore the DAGs with concepts relations). The concept nodes $\{C_1 \ldots C_K\}$ are of mixture-of-Gaussians type, and their parameters are learned after the network structure learning.

Table 1. Mean average precisions.

	Mean Average Precision	MAP with 100 images per query
Test set	13.8%	17.8%
Train set	14.7%	19.2%

3. EXPERIMENTS AND RESULTS

The algorithm was tested with the subset of COREL Stock imahes that was used by Duygulu et al. in [8] and evaluated with average precision and mean average precision. The low-level features we used are: Tamura, Gabor, and marginal HSV color, see [9]. To reduce the feature space dimensions we used PCA, ICA and the cross-entropy based algorithm [5]. Only concepts with more than 100 training examples were considered (36 total). Figure 2 presents the average precision of the tested concepts and the 5 best/worst concepts' average precisions. Table 1 presents the mean average precisions for the test and train set. The reported measures refer to the results using PCA feature selection.



sky	50.1%
tracks	46.5%
water	45.8%
cars	27.1%
jet	27.0%
…	
close-ups	4.8%
hills	4.2%
valley	4.2%
stone	4.0%
statue	2.9%

Figure 2. Average precision by concept.

4. DISCUSSION AND FUTURE WORK

Our aim is to detect salient concepts with a high degree of confidence (average precision) and stability across different datasets. From the evaluation results we conclude that it is possible to use unsupervised methods to extract patterns and label those patterns (or a combination) with a concept. However, the discovery of the optimal pattern-concept relation is difficult. Interesting facts could be observed: different feature selection methods returned very different average precisions for certain concepts; and supposedly easy concepts (e.g. hills, valley), present a very low average precision.

These measures show the validity of our approach. In the future we will (1) research the use of other algorithms besides K2 and finite-mixtures to learn the concepts' models; (2) study the effect of using features subspaces and smaller-granularity clusters; (3) and finally, test the framework with other datasets.

5. ACKNOWLEDGMENTS

This work was partially funded by the Portuguese Foundation for Science and Technology.

6. REFERENCES

[1] A. W. M. Smeulders, M. Worring, S. Santini, A. Gupta, and R. Jain, "Content-based image retrievals at the end of the early years," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, pp. 1349-1380, 2000.

[2] T. Westerveld and A. P. d. Vries, "Experimental result analysis for a generative probabilistic image retrieval model," presented at Proceedings SIGIR, Toronto, Canada, 2003.

[3] J. Jeon, V. Lavrenko, and R. Manmatha, "Automatic image annotation and retrieval using cross-media relevance models," presented at Proceedings of SIGIR, Toronto, Canada, 2003.

[4] T. Hastie, R. Tibshirani, and J. Friedman, *The elements of statistical learning: Data mining, inference and prediction*: Springer, 2003.

[5] D. Koller and M. Sahami, "Toward Optimal Feature Selection," presented at International Conference on Machine Learning (ICML), Bari, Italy, 1996.

[6] M. Figueiredo and A. K. Jain, "Unsupervised learning of finite mixture models," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, pp. 1-16, 2002.

[7] R. Neapolitan, *Learning Bayesian Networks*: Prentice Hall, 2003.

[8] P. Duygulu, K. Barnard, N. d. Freitas, and D. Forsyth, "Object recognition as machine translation: Learning a lexicon for a fixed image vocabulary," presented at European Conference on Computer Vision, 2002.

[9] D. Heesch, P. Howarth, J. Magalhães, A. May, M. Pickering, A. Yavlinsky, and S. Rüger, "Video retrieval using search and browsing," presented at TRECVID Workshop 2004, Gaithersburg, USA, 2004.