Incremental Multiple Classifier Active Learning for Concept Indexing in Images and Videos

Bahjat Safadi, Yubing Tong, Georges Quénot

Incremental Multiple Classifier Active Learning for Concept Indexing in Images and Videos

Bahjat Safadi, Yubing Tong and Georges Quénot

Laboratoire d'Informatique de Grenoble
BP 53 - 38041 Grenoble Cedex 9, France
{Bahjat.Safadi, Yubing.Tong, Georges.Quenot}@imag.fr

Abstract. Active learning with multiple classifiers has shown good performance for concept indexing in images or video shots in the case of highly imbalanced data. It involves however a large number of computations. In this paper, we propose a new incremental active learning algorithm based on multiple SVM for image and video annotation. The experimental result show that the best performance (MAP) is reached when 15-30% of the corpus is annotated and the new method can achieve almost the same precision while saving 50 to 63% of the computation time.

Keywords: Multimedia Indexing, Machine Learning, Active Learning and Incremental Learning

1 Introduction

Supervised learning consists in training a system from sets of positive and negative examples. The learning system may be composed of various types of feature extractors, classifiers and fusion modules. The performance of the systems depends a lot upon the implementation choices and details but it also strongly depends upon the size and quality of the training examples. While it is quite easy and inexpensive to get large amounts of raw data, it is usually very expensive to have them annotated because it involves human intervention for the judging of the "ground truth". While the volume of data that can be manually annotated is limited due to the cost of manual intervention, there remains the possibility to select the data samples that will be annotated so that their annotation is as useful as possible ([1]). *Active Learning* is a special case of machine learning which has been used to improve query performance in image retrieval systems. The objective of Active learning is to maximize the expected information from the query as a result of user feedback in order to minimize the total number needed for the search. This can be summarized as following, from relevance feedback, a user subjectively labels the retrieved images as Positive or Negative, and these labelled images are used to train a classifier that performs a bi-class classification on the image database. Those images with the higher scores or probability values, with respect to the positive image class, are retrieved as the most informative samples to be labelled by the user. However, in very large databases, the learning performance is often restricted due to a very small number of available labelled samples to a given concept, because labelling concepts in too many images or videos is a very hard task, and a user is unwilling to label too many retrieved images for relevance feedback. Classifiers (such as Support Vector Machines) based active learning has been proposed [3, 9, 12, 13, 16] to handle this problem by maximizing the learning efficiency while minimizing the required number of labelled image samples for training process.

Recently, some researches like in [12, 17, 18] have shown the effectiveness of using the multi-learners approach to handle the problem of the imbalances between the major and minor classes in the very large scale databases. This problem is very common in concept indexing in images and videos since the most target concepts are very sparse. For instances, their average frequency in the TRECVID evaluation campaigns [15] for example, is less than 1%. This imbalance is a serious problem for classical supervised learning methods. An alternative approach to solve this problem is to sub-sampling the major class [4] (the negative samples); this sub-sampling can be done by considering all the positive samples in the training set and selecting randomly a comparable number of negative to positive samples. This sub-sampling might leads to loss of information, due to the fact that it ignores a lot of information from the non chosen negative samples, hence the multi-learner has the ability to balance the loss of information related to this sub-sampling by making several selections on this class and fusing the outputs of different classifiers built from these subsets. In [12],our previous work , we showed that combining between the Active Learning with multi learners approaches significantly increases the effectiveness of the Active Learning, but it makes it very slower comparing to a mono-learner approach. This makes a very big challenge in the task of automatically image annotation, which mostly is directed by learning from user's feedback.

Since during the iteration of active learning and multi-learners (here multiple SVMs are used), new labelled samples will always be added to the training set for next iteration, each iteration involves previous training information and the new untrained samples. The calculation time will be saved if we can re-use the previous information and learn the incremental information derived from new samples. So it is natural to adopt incremental learning for this case.

Some incremental learning focus on how to choose the informative samples data from all the incremental one or retire some samples from previous set [14, 20, 21]. Those methods needs to check KKT conditions of SVM quadratic optimization problem for every sample which also means much calculation. [6, 7] also consider the calculation problem in active learning and multi-learner. An early stopping method is proposed to achieve faster convergence of active learning by counting the number of support vectors derived from previous training in [5]. If the number of support vectors stabilizes, it means that all possible SVs have been selected by active learning method. This method may lose some useful information since the number of SVs may still change after several stable values and the stability of SVs is not clearly defined. Our previous multi-learner active learning research also shows that for every learner in iteration only a part of the samples used for training [12, 13]. But many classifiers are still needed to train during iteration. Although some samples in previous step have been well trained but they are never used in the following step. Some researchers have tried some incremental methods used in SVM training. [11] proposed an incremental learning of SVM, that The support vectors from previous training set will be used involved in the new SVM optimization problem with different weights on them. This method can work for balanced data. But for our extremely imbalanced data, the weight is so rough that the final hyperplane deviated much from the ideal one. Furthermore this method also needs to train the previous support vector, then no time can be saved. In [19] authors proposed an incremental learning of SVM by classifier combining. Multiple SVM are used and each output posterior probability information. For the example incremental learning, the training set can be divided into several learning sequence or incremental batch. Cross validation was made on every batch and classifier is also trained, then the output of each classifier predicting on testing sample can be combined to get the average posterior probability. Here every batch works independently without using the information from previous training.

In this paper, we overcome the problem of processing time for the Active learning with multiple-learners by proposing a robust Incremental algorithm for the Active learning based multiple-SVMs and we show that we can save about 50-63% of the processing time (depending on the used descriptor and the negative to positive ratio), while the system performance was not significantly changed in all cases; our experiments were conducted on the TRECVID 2007 and 2008 collections.

The outline of the paper continues as follows: the combination between the multiple classifier and active learning approach is discussed in section 2. We present our new incremental learning algorithm in introduced in section 3; section 4 describes the experimental results including the description of the used data collection and the descriptors, while Section 5 presents concluding remarks.

2 Active learning with multiple classifiers

Active learning has been adopted to solve problems related to unlabelled training data. For imbalanced data, many papers have shown the possibility to balance the loss of information related to sub-sampling of the negative class by making several selections on this class set and fusing the outputs of different classifiers built from these subsets. This leads to what we call the Multi-learners approach. The active learning algorithm with multiple classifiers is detailed in Algorithm 1 which is a classical active learning algorithm in which we have replaced the single classifier by a set of elementary classifiers. For implementation purposes, the elementary learning algorithm A is split into two parts: Train and Predict. A global parameter, mono-learner, can force the classical active learning mode with a single classifier. At each iteration i, the development set S is split into two parts: L_i, labeled samples and U_i, unlabelled samples. A global parameter f_{pos} defines the ratio between the negative and positive samples in all learners and for all iterations. This defines the number of negative samples for each learner at iteration i. In the multi-learner approach, the number of learners is computed so that each negative sample appears in average a given number of times (usually once) in the different subsets T_j. The T_j contains all positive samples and a randomly chosen subset of negative samples. Classifiers C_j are then trained on the T_j with associated labels and applied to U_i the unlabelled set for the selection of the next samples to annotate. Predictions from the elementary classifiers are then merged in both cases for producing a single prediction score per sample. The predictions on the U_i set are used by Q the selection (or querying) function to produce a sorted list of the next samples to be anno-

Algorithm 1 Multiple Classifier Active Learning Algorithm

S: all data samples.
L_i, U_i: labelled and unlabelled subsets of S.
A=(Train, Predict): the elementary learning algorithm.
Q: the selection (or querying) function.
$nl(k)$: number of learners at iteration k.
Initialize L_i (e.g. 10 positives & 20 negatives).
while $S \setminus L_i \neq \emptyset$ **do**
  if mono-learner **then**
    $nl(k) = 1$
  else
    $nl(k) =$ Calculate the number of Learners
  end if
  for all $j \in [1..nl(k)]$ **do**
    Select subset T_j from L_i for training
    $C_j \leftarrow$ Train(T_j)
    $P_{un}^j \leftarrow$ Predict(U_i,C_j)
  end for
  $P_{un} \leftarrow$Fuse(P_{un}^j)
  Apply Q on P_{un} and select $\tilde{x} \in U_i$ samples.
  $\tilde{y}$ = Label $\tilde{x}$
  $L_{i+1} \leftarrow L_i \cup (\tilde{x};\tilde{y})$
  $U_{i+1} \leftarrow U_i \setminus \tilde{x}$
end while

tated. From the top of this list, a $\tilde{x}$ set is selected for annotation. The $\tilde{x}$ set is then added with the associated set of labels $\tilde{y}$ to the L_i set to produce the L_{i+1} set and it is also removed from the U_i set to produce the U_{i+1} set. The global algorithm is determined by the A=(Train, Predict) elementary learning algorithm (e.g. SVM) and by Q the selection (or querying) function implementing the active learning strategy (e.g. relevance or uncertainty sampling). It is also determined by some global parameters like the ratio between the number of negative and positive samples, by the cold start problem, by the fusion function used to fuse the outputs of the classifiers and by the way we choose the number of new samples to be integrated at each iteration. For our evaluation experiments we show the system performance by calculating the Mean average precision on the test set at each step.

3 The proposed incremental method

As described in section 2, algorithm 1 can be used to handle the class imbalance problem. Even though it gave good results [12, 13, 17], it is still very slow because at each iteration it generates a lot of learners when the dataset is highly imbalanced. In this section we propose an incremental method to reduce the number of learners that need to be trained at each iteration. Let $nl(k)$ be the number of learners needed at step k, $nm[k]$ to be the minimum number of learners to be changed at step k. At each iteration the actual number of learners which should be trained is equal to $nl(k) - (nl(k-1) - mn[k])$. After training these learners we merge their results with the results obtained from $nl(k-1) - mn[k]$ learners from the previous steps, so that at each iteration we keep the $nl(k)$ of this iteration but we only train part of them as shown in table 1 where rm and add indicate respectively the number of learners should be removed and added at each iteration.

At step k:
$nl(k)$: number of learners at step k
$nl(k-1)$: learners trained from the previous steps
$nm[k]$: minimum number of learners to be changed
The number of learners to be removed from the previous step:
$if(nl(k) >= nl(k-1))\ \ rm = nm$
$if(nl(k) < nl(k-1))\ \ rm = nl(k-1) - nl(k) + nm$
The number of learners to be added by:
$if(nl(k) <= nl(k-1))\ \ add = nm$
$if(nl(k) > nl(k-1))\ \ add = nl(k) - nl(k-1) + nm$

Table 1. The conditions of removing and adding learners

In figure 4 we show one step of how the proposal algorithm works. At each iteration the algorithm should remove the learners with the minimum number of positive samples (normally the learners taken from the oldest iterations), and train some new learners, each learner should be trained on a subset that consists of all the positive samples and of a comparable number of negative samples randomly selected from the training set. At the end it applies the fusion function on the results obtained from the considered learners to give the final score of each unlabelled sample.

In our experiments, we fixed the minimum and maximum values of nm to be $\{1,10\}$ respectively, and $nm[k] = 20\% nl(k)$.

4 Experiments

We have evaluated the Multiple Classifiers Active Learning and the proposed incremental methods in a variety of contexts. It has been applied using four types of image descriptors using the SVM with RBF kernel as classifier and the relevance sampling strategy for active learning. We also used the harmonic Mean function to fuse the results of the multi learners. The cold start problem was not really explored; a random set of 10 positive and 20 negative samples was used. The global parameters like the f_{pos} ratio were taken from our previous work [13]. For the number of samples to be added at each iteration, we chose a variable step size since we observed in previous experiments that having small steps in the beginning of the active learning process is better for the speed of performance improvement. In practice, we used a geometric scale

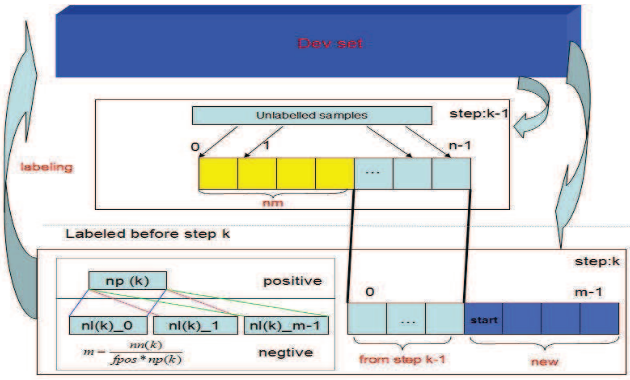

Fig. 1. The framework of the proposed incremental method

with 40 steps. The evaluations were conducted using the TRECVID 2007 and 2008 test collections and protocols.

4.1 TRECVID 2007 and 2008 collections

The evaluation was conducted on TRECVID-2008 concepts annotated on the TRECVID 2007 and 2008 collections where 20 concepts were evaluated. The training and evaluation were done respectively on the development and test sets of the two collections, TRECVID 2007 collection contains 21532 video shots as a training set and 22084 shots as test set, while TREC2008 contains 43616 video shots as a training set and 35766 shots as a test set. In the two collections, the training sets are fully annotated and nothing remains to be annotated which makes the use of the Active learning not relevant, but such large fully annotated sets constitute opportunities to simulate, evaluate and compare strategies and methods in active learning without the need of involving a user, as the simulated active learning [2]. In our experiments active learning methods are executed as if very few annotations are available in the training set. Then, each time a human annotation is needed, the corresponding subset of the full annotation is made available to the active learner.

4.2 Image representation

Concepts and Images can be represented by their vector descriptors or features. Many descriptors could be used to represent a specific concept in an image, finding the best descriptor to represent a concept in an image is still a big challenge and a wide area of research. For evaluations we used four descriptors of different types and sizes that have been produced by various partners of the IRIM project of the GDR ISIS [10] .

- LIG_hg104: early fusion with normalization of an RGB histogram $4 \times 4 \times 4$ and a Gabor transformation (8 orientations and 5 scales), $64 + 40 = 104$ dimensions.
- CEALIST_global_tlep: early fusion of local descriptors of texture and of an RGB color histogram, $512 + 64 = 576$ dimensions.

- ETIS_global_qwm1x3x256: 3 histograms of 3 vertical bands of visual descriptors, standard Quaternion wavelet coefficients at three scales, $3 \times 256 = 768$ dimensions.
- LEAR_bow_sift_1000: histogram of local visual descriptors, SIFT "classic" [8], 1000 dimensions.

4.3 Optimal negative to positive ratios

Table 2 shows the optimal values for the f_{pos} global parameter on the development set for single- and multiple-learner versions SVM-RBF with the four considered descriptors. Optimization was done while taking all the development set of TRECVID2007, in the Multiple-learner versions the results of the classifiers are fused by the harmonic mean function. These optimal values are higher for the single learner than for the multiple learner case. This was expected since the multiple-learner has another way to take into account more negative samples in total.

Descriptor	Single	Multi
LIG_hg104	4	2
CEALIST_global_tlep	8	4
ETIS_global_qwm	4	3
LEAR_bow_sift	8	4

Table 2. Optimal values of the ratio between the numbers of negative to the positive samples for four descriptors.

4.4 The active learning steps

In our evaluation, we used totally 40 steps for the active learning algorithm, considering the geometric scale function with the following formula:

$$S_k = S_0 \times \left(\frac{N}{S_0}\right)^{k/K}$$

where N is the total size of the development set, S_0 is the size of the training set at the cold-start(30 samples), K is the total number of steps and k is the current step. At each step (or iteration) the algorithm calculates the S_k to be the size of the new training set and it chooses new samples to be labelled with size equal to $S_k - S_{k-1}$.

4.5 Active learning effectiveness

Figures 2 and 3 compare the effectiveness of the three methods (the single and Multi-learner and the Incremental) using the four descriptors and the relevance sampling strategy. The performance of the Single-learner is also shown as baseline. For the multiple-learner and the incremental experiments, the fusion by harmonic mean has been used. These plots show the evolution of the indexing performance of the test sets measured

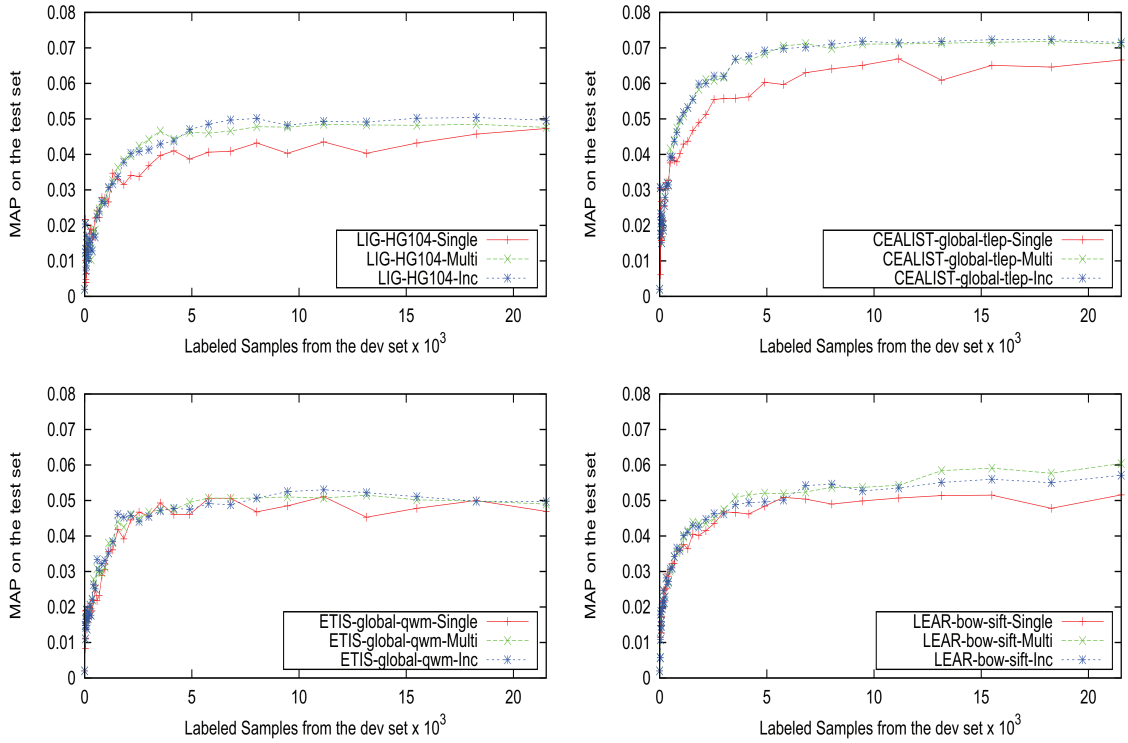

Fig. 2. The Map results on the TRECVID 2007 test collection evaluated on the four descriptors, each one of the plots shows the results using the Single-learner (in red), the Multi-learner (in green) and the Incremental (in blue).

by the Mean Average Precision (MAP) metric with the number of annotated samples at each step (totally 40 steps). For analysing the plots we consider that The faster it grows and the higher performance it achieves , especially in the beginning, the better.

As the plots shown in figure 2 and 3, the proposed incremental algorithm has almost achieved the same performance as that of multi-learner. Both of them are significantly higher and faster to reach the highest value than single learner. With our incremental learning method, the highest performance can be reached with training by only 15-30% samples which means labelling 15-30% samples instead of all the samples can still get the same result (in MAP).

We considered the index of G_{a-b} to be the performance measure between two active learning curves(a and b), this measure was calculated simply as following: $G_{a-b} = (A_a - A_b)/A_b$ where A_i indicts the area of the curve i, to calculate this gain we first normalize the curves in each plot, then we calculate the area using the following formula:

$$A = \frac{1}{2}\left|\sum_{i=0}^{n+1} x_i \times y_{i+1} - y_i \times x_{i+1}\right|$$

Where n is the total number of iterations (x_i, y_i) indicates the (number of the annotated samples and the MAP value) at iteration i, and $(x_{n+1}, y_{n+1}) = (x_0, y_0)$. Table 3 show the gain when using the incremental method compared to both the single and

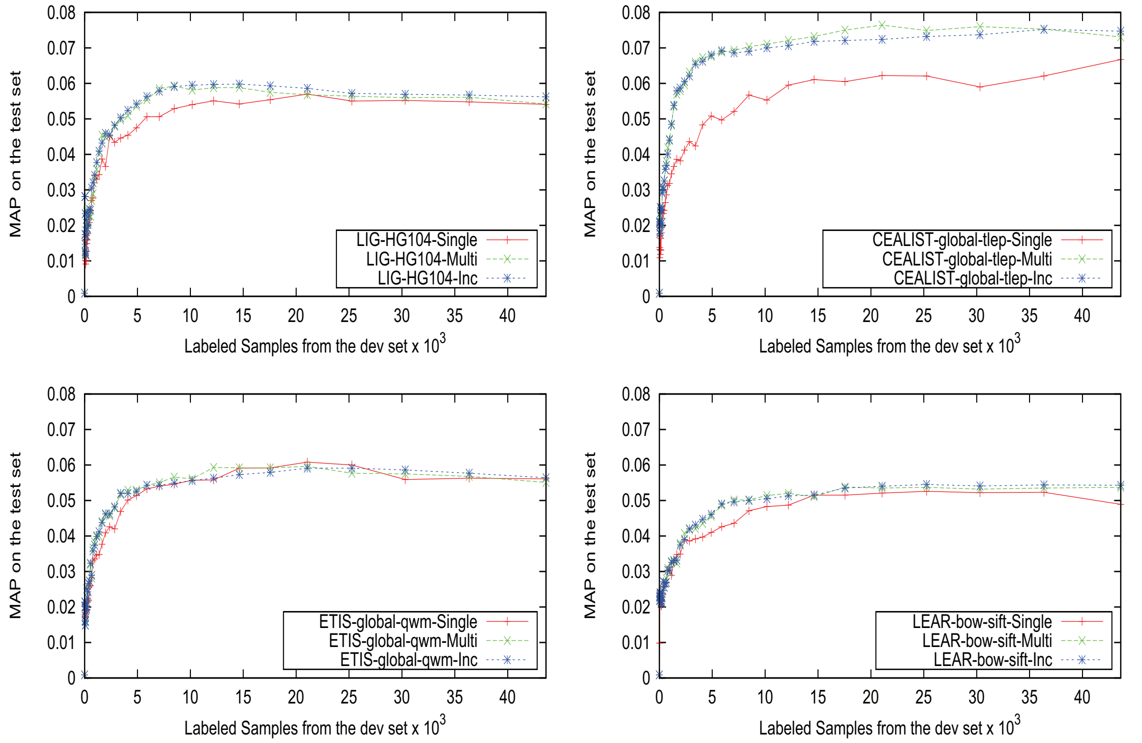

Fig. 3. The Map results on the TRECVID 2008 test collection evaluated on the four descriptors, each one of the plots shows the results using the Single-learner (in red), the Multi-learner (in green) and the Incremental (in blue).

multi-learner methods with the two collections that considered in our experiments, as we can see that the gain is much higher and significant when using our incremental method compared to the single-learner G_{I-S} while it is very small compared to the multi-learner G_{I-M}.

Descriptor	Trecvid 2007		Trecvid 2008	
	$G_{I-S}(\%)$	$G_{I-M}(\%)$	$G_{I-S}(\%)$	$G_{I-M}(\%)$
LIG_hg104	14.77	2.34	6.50	1.83
CEALIST_global_tlep	12.84	0.62	22.42	-1.70
ETIS_global_qwm	4.76	0.73	1.20	0.02
LEAR_bow_sift	8.04	-3.16	5.22	0.75

Table 3. The gain of the system performance between the proposed incremental to the single- and the multi-learners, with the four descriptors evaluated on TRECVID 2007 and 2008.

4.6 Execution times

Table 4 gives the total execution times for the whole active learning process (40 iterations) on all 20 concepts on each experiment collection, per method and per descriptor, using the relevance strategy. As we can see that Single learner is faster than multiple learner and incremental methods. But considering the performance of single learner described in the above section, the performance of single learner is much lower than that of multi-learner. Compared with multiple learners, the new proposed incremental method has saved nearly 48-66% time without losing any performance.

Descriptor	Trecvid 2007				Trecvid 2008			
	Single	Multi	Inc	G	Single	Multi	Inc	G
LIG_hg104	1.40	20.63	7.64	66%	4.80	59.54	23.34	60%
CEALIST_global_tlep	23.90	115.02	64.17	52%	96.56	395.45	204.9	48%
ETIS_global_qwm	13.40	142.97	64.10	55%	45.67	460.60	212.3	54%
LEAR_bow_sift	43.42	162.18	79.16	52%	181.00	592.10	300.6	49%

Table 4. Processing time table for the two evaluated collections: TRECVID 2007 and 2008, with G that indicates the gain of time using our incremental method compared to the multi-learners.

5 Conclusion

Active learning with multiple classifiers has shown good performance for concept indexing in images or video shots in the case of highly imbalanced data. It involves however a large number of computations. In this paper, we propose a new incremental active learning algorithm based on multiple SVM for image and video annotation. The experimental result show that the best performance (MAP) is reached when 15-30% of the corpus is annotated and the new method can achieve almost the same precision while saving 50 to 63% of the computation time.

Acknowledgements

This work was partly realized as part of the Quaero Program funded by OSEO, French State agency for innovation.

References

1. D. Angluin. Queries and concept learning. Machine Learning, 2:319–342, 1988.
2. S. Ayache and G. Quénot. Evaluation of active learning strategies for video indexing. Image Commun., 22(7-8):692–704, 2007.
3. S. Ayache, G. Queńot and J. Gensel. Image and video indexing using networks of operators. EURASIP Journal on Image and Video Processing, 2007(4):1–13, 2007.
4. C. M. Bishop. Pattern Recognition and Machine Learning (Information Science and Statistics). Springer, 1 edition, October 2007.
5. A. Bordes, S. Ertekin, J. Weston and L. Bottou. Fast kernel classifiers with online and active learning. JOURNAL OF MACHINE LEARNING RESEARCH, 6:1579–1619, 2005.

6. S. Ertekin, J. Huang and C.L. Giles. Active learning for class imbalance problem. In SIGIR 07: Proceedings of the 30th annual international ACM SIGIR conference on Research and development in information retrieval, pages 823–824, New York, NY, USA, 2007. ACM.
7. S. E. Ertekin, J. Huang, L. Bottou, and C. L. Giles. Learning on the border: Active learning in imbalanced data classification. In In Proc. ACM Conf. on Information and Knowledge Management (CIKM 07, 2007).
8. D. Lowe. Distinctive image features from scale-invariant key points. International Journal of Computer Vision, 60(2):91–110, 2004.
9. M.R. Naphade and J.R. Smith. On the detection of semantic concepts at trecvid. In MULTIMEDIA 04: Proceedings of the 12th annual ACM international conference on Multimedia, pages 660–667, New York, NY, USA, 2004. ACM Press.
10. G. Quénot, B. Delezoide, H. le Borgne, P.-A. Moellic, D. Gorisse, F. Precioso, F. Wang, B. Merialdo, P. Gosselin, L. Granjon, D. Pellerin, M. Rombaut, H. Bredin, L. Koenig, H. Lachambre, E. E. Khoury, B. Mansencal, J. Benois-Pineau, H. Jégou, S. Ayache, B. Safadi, J. Fabrizio, M. Cord, H. Glotin, Z. Zhao, E. Dumont, and B. Augereau. Irim at trecvid 2009: High level feature extraction. In TREC2009 notebook, 16-17 Nov 2009.
11. S. Ruping. Incremental learning with support vector machines. In ICDM 01: Proceedings of the 2001 IEEE International Conference on Data Mining, pages 641–642, Washington, DC, USA, 2001. IEEE Computer Society.
12. B. Safadi and G. Quénot. Active learning with multiple classifiers for multimedia indexing. In CBMI, Grenoble, France, June 2010.
13. B. Safadi and G. Que not. Evaluations of multi-learners approaches for concepts indexing in video documents. In RIAO, Paris, France, Apr 2010.
14. A. Shilton, M. Palaniswami, S. Member, D. Ralph, A. C. Tsoi, and S. Member. Incremental training of support vector machines. IEEE Transactions on Neural Networks, 16:114–131, 2005.
15. A. F. Smeaton, P. Over, and W. Kraaij. Evaluation campaigns and trecvid. In MIR 06: Proceedings of the 8th ACM International Workshop on Multimedia Information Retrieval, pages 321–330, New York, NY, USA, 2006. ACM Press.
16. C. G. M. Snoek, M. Worring, and A. G. Hauptmann. Learning rich semantics from news video archives by style analysis. ACM Trans. Multimedia Comput. Commun. Appl., 2(2):91–108, 2006.
17. M. A. Tahir, J. Kittler, K. Mikolajczyk, and F. Yan. A multiple expert approach to the class imbalance problem using inverse random under sampling. In MCS 09: Proceedings of the 8th International Workshop on Multiple Classifier Systems, pages 82–91, Berlin, Heidelberg, 2009. Springer-Verlag.
18. M. A. Tahir, J. Kittler, F. Yan, and K. Mikolajczyk. Concept learning for image and video retrieval: The inverse random under sampling approach. In Eusipco 2009, 17th European Signal Processing Conference, August 24-28 2009.
19. Y.-M. Wen and B.-L. Lu. Incremental learning of support vector machines by classifier combining. In PAKDD-07: Proceedings of the 11th Pacific-Asia conference on Advances in knowledge discovery and data mining, pages 904–911, Berlin, Heidelberg, 2007. Springer-Verlag.
20. C. Wu, X. Wang, D. Bai, and H. Zhang. Fast incremental learning algorithm of svm on kkt conditions. In FSKD 09: Proceedings of the 2009 Sixth International Conference on Fuzzy Systems and Knowledge Discovery, pages 551–554, Washington, DC, USA, 2009. IEEE Computer Society.
21. Y.Zhang,X.Wang,andJ.Zhai.Afast support vector machine classification algorithm based on karush-kuhn-tucker conditions. In RSFDGrC 09: Proceedings of the 12th International Conference on Rough Sets, Fuzzy Sets, Data Mining and Granular Computing, pages 382–389, Berlin, Heidelberg, 2009. Springer-Verlag.