Improving Active Learning with Sharp Data Reduction

Priscila T. M. Saito[†], Pedro J. de Rezende[†], Alexandre X. Falcão[†],
Celso T. N. Suzuki[†], Jancarlo F. Gomes[†‡]
[†]Institute of Computing, University of Campinas - UNICAMP, Campinas, SP, Brazil
[‡]Institute of Biology, University of Campinas - UNICAMP, Campinas, SP, Brazil
{maeda, rezende, afalcao, celso.suzuki, jgomes}@ic.unicamp.br

ABSTRACT

Statistical analysis and pattern recognition have become a daunting endeavour in face of the enormous amount of information in datasets that have continually been made available. In view of the infeasibility of complete manual annotation, one seeks active learning methods for data organization, selection and prioritization that could help the user to label the samples. These methods, however, classify and reorganize the entire dataset at each iteration, and as the datasets grow, they become blatantly inefficient from the user's point of view. In this work, we propose an active learning paradigm which considerably reduces the non-annotated dataset into a small set of relevant samples for learning. During active learning, random samples are selected from this small learning set and the user annotates only the misclassified ones. A training set with new labelled samples increases at each iteration and improves the classifier for the next one. When the user is satisfied, the classifier can be used to annotate the rest of the dataset. To illustrate the effectiveness of this paradigm, we developed an instance based on the optimum path forest (OPF) classifier, while relying on clustering and classification for the learning process. By using this method, we were able to iteratively generate classifiers that improve quickly, to require few iterations, and to attain high accuracy while keeping user involvement to a minimum. We also show that the method provides better accuracies on unseen test sets with less user involvement than a baseline approach based on the OPF classifier and random selection of training samples from the entire dataset.

Keywords: Pattern Recognition, Machine Learning, Active Learning, Semi-Automatic Dataset Annotation, Data Mining, Optimum-Path Forest Classifiers.

1 INTRODUCTION

The amount of available information has been increasing due to the advances of computing and data acquisition technologies, resulting in large datasets. Handling and analysing such increasing volume of information have become humanly infeasible and highly susceptible to errors, since it is extremely time consuming and wearisome. Hence, there is an increasing demand for the development of effective and efficient ways to annotate these datasets.

Active learning techniques have been explored and reasonably successful. However, these methods fall in a single paradigm which requires, at each iteration, the classification of the entire dataset under annotation, followed by the organization of all these samples according to some criterion, in order to select the most informative samples to be used for training the classifier. These phases are highly interdependent and, for large datasets, performing them at each iteration is very inefficient or even computationally infeasible.

In this paper, to overcome the aforementioned problems, we propose a new active learning paradigm which is verifiably effective and more efficient in practice, when dealing with large datasets, than those based on the current state of the art. The proposed paradigm relies on a significant reduction in the dataset size to create a small representative set of samples, for the learning process. By constructing the first instance of the classifier based on the knowledge of as many classes as possible, as well as incorporating the best samples at each iteration, subsequent selection and classification phases are much more efficacious. This approach differs from the traditional active learning methods, in which *all* samples in the database have to be classified and re-organized at each iteration.

Being a paradigm, it can be implemented using different strategies. This paper also presents an instantiation (Cluster-OPF-Rand) which has been developed to illustrate the effectiveness of this paradigm. It is based on the Optimum Path Forest (OPF) classifier, while relying on clustering and classification for the learning process. Cluster-OPF-Rand prevents the user from having to annotate a large (and usually wasteful) number of training samples. Moreover, it prevents poor selection of sam-

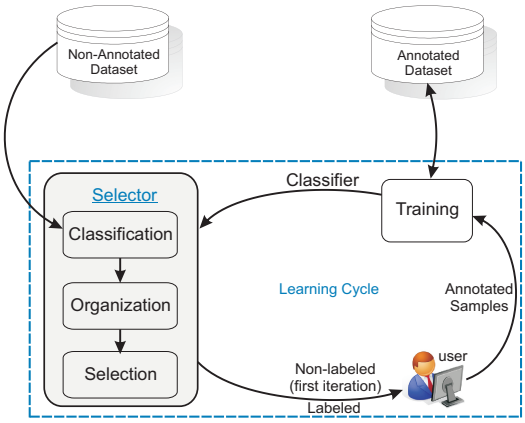


Figure 1: Pipeline of the traditional active learning paradigm.

ples from a large learning set, since this set is reduced so as to contain essentially the most representative samples. After this reduction, the proposed paradigm enables the organization of the learning samples to occur beforehand (and only once). In this particular implementation, the organization of the reduced set occurs in a randomized fashion.

The experiments performed on three datasets show that Cluster-OPF-Rand is interactively and iteratively efficient, in addition to providing high accuracies earlier. That is to say, the number of learning iterations is significantly reduced with better accuracies, while requiring the annotation of only a small number of samples, when compared to a baseline approach using the OPF classifier and random selection of training samples from the entire dataset. The results also showed impressive reductions of over 90% in user effort, at the same time providing accuracies of over 97%.

The remainder of this paper is structured as follows. Section 2 summarizes the major active learning techniques presented in the literature. Section 3 presents the clustering approach based on optimum-path forest used. Section 4 details the active learning paradigm and the reduced method proposed. Section 5 discusses the experiments and the accomplished results. Finally, Section 6 presents the conclusions and future work.

2 BACKGROUND AND TECHNIQUES

Recent works in active learning have yielded a variety of heuristics, which are designed mostly for binary classification and are applicable primarily to classifiers such as Artificial Neural Network (ANN), Support Vector Machine (SVM), k-Nearest Neighbour (k-NN) and Optimum-Path Forest (OPF).

In active learning techniques, the key idea relies on the strategy used to select the most informative samples such that they allow for the achievement of greater accuracies with fewer training labels annotated by the user. Much effort has been placed in investigating strategies for active learning. However, it is focused on methods that classify all samples in the database, then organize these samples according to certain criteria and subsequently select and display the most informative samples to be annotated by the user, at each learning iteration. For large databases, these complete phases, at each learning iteration, are very inefficient or even impractical to be done computationally.

Figure 1 illustrates the execution pipeline of the traditional active learning paradigm presented in prior literature. This paradigm is comprised of a learning algorithm and a selector. The selector consists of three modules (classification, organization and selection) that are highly interdependent. At each iteration cycle, the system presents to the user a set of samples that consists of either non-labelled samples (from the entire database, in the first iteration) or labelled ones (obtained through the classifier), all chosen by the selector. As these samples are annotated by the user, they are included in the training set to retrain the classifier for the next cycle.

Besides the aforementioned inefficiency, most of the existing research in the traditional active learning paradigm has focused on binary classification. Relatively few approaches [12, 20, 9, 16, 11, 10] have been proposed for multiclass active learning and are typically based on extensions of predominantly binary active learning methods to the multiclass scenario.

In the ANN literature, although several works [4, 1, 7] have explored the use of active learning in the context of efficient network training, this approach shows the disadvantage of being computationally expensive.

Alternatively, SVM has been used in [19, 18], under the assumption that the samples closest to the separating hyperplane are the most informative ones. During the iterations of relevance feedback, the method finds the optimal hyperplane separating relevant and irrelevant samples and presents to the user the samples closest to this hyperplane. This hyperplane is adjusted throughout the iterations, and after the last one, the method presents the most relevant samples as being the farthest ones to

the hyperplane. Extensions to the multiclass scenario are typically based on extensions of binary classification using pairwise comparisons or 1-vs-all strategy.

In contrast, [10] introduced a probabilistic variant of k-NN. Although, this variant was designed specifically for multiclass problems, it involves learning a certain number of parameters. Moreover, the performance of the method is dependent on the similarity measure used.

A strategy, similar to the one presented in [18], was proposed in [6], using a faster and more effective classifier based on Optimum-Path Forest (OPF). They developed greedy (GOPF) [5] and planned (POPF) [6] active learning strategies for CBIR systems. For a given set of relevant and irrelevant samples, the method computes an optimum-path forest using samples from the query set for training the classifier.

Optimum-Path Forest (OPF) is a framework for developing pattern classifiers (supervised, semi-supervised or unsupervised) which defines how the samples are connected by an adjacency relation that gives rise to a graph, and how to measure the connectivity (the cost of a path in the graph generated by the adjacency) between them by means of a function that gives rise to an optimum path forest.

The supervised and the unsupervised classifiers were described in [14, 17], respectively. Both learning approaches are fast and robust for large datasets [13, 2]. In addition, the classes/clusters may present arbitrary shapes and have some degree of overlapping. Classifiers based on OPF have been widely used in several applications and have demonstrated that OPF-based classifiers can be more effective and much faster than ANN and SVM based ones [14].

The following Section details the OPF based clustering approach.

3 CLUSTERING BY OPTIMUM-PATH FOREST

The data reduction approach we implemented is based on clustering by Optimum-Path Forest (OPF) [17]. This is a non-parametric approach which estimates the number of natural groups in a dataset as the number of maxima of its probability density function (pdf). In this approach, each maximum of the pdf will define a cluster as an optimum-path tree rooted at that maximum. It can handle plateaux of maximum, by electing a single root (one prototype per maximum), groups with arbitrary shapes, and some overlapping among clusters.

In this unsupervised learning algorithm, an unlabelled training set is interpreted as a graph whose nodes are samples (images, in this paper) and each node is connected with its k-closest neighbours in the feature space to form directed arcs. The pdf value at each node is estimated from the distance between adjacent samples, and a connectivity (path-cost) function is designed such that the maximization of a connectivity map defines an optimum-path forest rooted at the maxima of the pdf. In this forest, each cluster is one optimum-path tree rooted at one maximum (prototype). The pdf estimation also requires multiple applications of the algorithm for different values of k in order to select the best clustering result as the one that produces a minimum normalized cut in the k-NN graph. The clusters are found by ordered label propagation from each maximum, as opposed to the mean-shift algorithm of [3] which searches for the closest maximum by following the direction of the gradient of the pdf — a strategy that does not guarantee the assignment of a single label per maximum, and presents problems on the plateaux of the pdf.

In order to handle large datasets, this approach estimates the pdf from random samples and fast propagates the group labels to the remaining samples of the dataset. The best k for pdf estimation is found by optimization, but its search interval [1, kmax] may produce different numbers of groups. The parameter kmax represents an observation scale for the dataset. If kmax is too high, it means that we are looking at the dataset from infinity and so, the result will be a single cluster. As we approximate the dataset (reducing the value of kmax), the number of clusters increases up to some high number for kmax = 1. Still, the number of possible solutions is low, because the method produces an identical number of clusters for several values of kmax. This shows the robustness of the method in finding natural groups in the dataset for distinct observation scales. In this work, we chose kmax so as to obtain a number of groups higher than the number of classes known. Note that, we do not use any knowledge on the classes of samples, but we assume that we know how many classes are present in the dataset.

4 PROPOSED PARADIGM

We propose a new paradigm for active learning in order to select, more efficiently and effectively, a small number of the most representative samples for training a classifier. The execution pipeline of the proposed paradigm is illustrated in Figure 2.

In the proposed paradigm, a classifier instance is generated at each iteration. After retraining the classifier (a process that relies on user annotations), the selector displays the most informative samples to the user. As the classifier improves, the user is required to correct fewer misclassified samples and progressively develops a sense of when the learning process has reached a satisfactory state.

Active learning methods presented in the literature differ from one another in their learning algorithms and in the selection strategies employed. The main difference

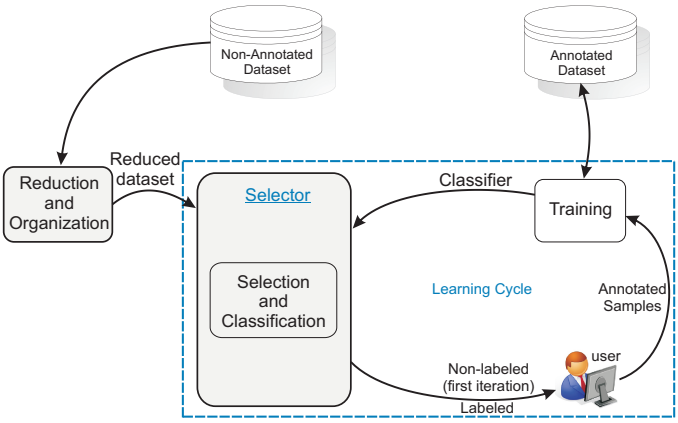


Figure 2: Pipeline of the proposed active learning paradigm.

between the proposed paradigm and previously proposed ones lies within the selector. Traditional methods make use of three modules that correspond to classification, organization and selection of samples (Figure 1). In these methods, the selection criterion is based solely on a classifier that is not yet reliable. When the classification accuracy is still low, the organization phase becomes useless, since when samples are classified, informative samples may not be selected to participate in the organization phase and therefore they will not be shown to the user.

The proposed paradigm is based on a priori data reduction and organization of the reduced dataset. It focuses on reversing the process adopted by traditional paradigms where an classification phase occurs before the organization phase. In the proposed paradigm, the selector consists of only one module of selection and classification. A major advantage presented by the proposed paradigm is that the reduction and organization of samples can be performed only once, unlike traditional methods.

Thus, the selector becomes faster, especially considering large databases, since the improvement of the classifier at each iteration does not require rearranging all samples; only the selection and classification phases are required. Moreover, a remarkably faster selection phase is completed by the choice of a small subset of samples and the classification of only these.

The strategy to be developed in order to select the most informative samples itself occurs as preprocessing in the module of reduction and organization (Figure 2). This strategy should not be based on a classifier, because it is still unreliable, but rather based on an absolute criterion previously established (for instance, exploring the organization of the data in the feature space). The classification phase is performed a posteriori, supporting the choice of the most informative samples by the selector, which follows a predetermined order in the reduction and organization module. In this module, different methods can be applied in our paradigm. In Subsection 4.1, we develop and present an effective method for the learning process.

4.1 Instantiation of the proposed paradigm

As it was mentioned, any method can be incorporated into the proposed paradigm in order to reduce the learning set and later to organize the reduced one. In this section, we present an effective method called Cluster-OPF-Rand. Figure 3 illustrates an example of the pipeline of Cluster-OPF-Rand.

The proposed method is divided into two modules: (1) reduction and organization, (2) selection and classification. The reduction and organization module is comprised of two steps: clustering and reduction of the data (steps 1 and 2 of Figure 3, respectively). The selection and classification module choose and label (steps 3 and 4 of Figure 3, respectively) the most informative samples of the reduced set chosen in a randomized fashion. Each sample is represented by a pair (id, lbl), where id corresponds to the identifier of the sample and lbl corresponds to the label given by the classifier. Note that it does not classify *all* samples in the dataset, but only the selected subset.

Initially, clusters are computed in order to obtain samples of all classes, as described in Section 3. One or more clusters represent samples of all classes in the non-labelled set, so that each cluster comprises mostly samples of a single class. Then, their roots (highlighted after step 1) cover samples of all classes and are defined as an initial training set for manual annotation. This is fundamental to be able to train the classifier with samples of all classes, since the first iteration. This classifier should be as good as possible because it is used in the classification of samples, providing an initial labelling, in which the user is not required to annotate all samples shown but only to correct a small number of misclassified ones.

Besides knowing which samples are roots of clusters, it is possible to identify those that are boundary sam-

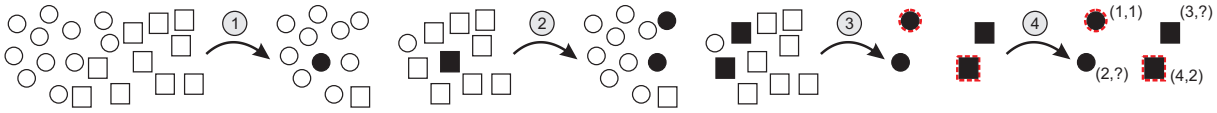


Figure 3: An example of pipeline of the proposed method.

ples between different clusters. A sample s is considered a boundary sample if there exists, among its k-NN adjacent samples, at least one whose label (given by the clustering) is different from that of s. The cluster boundary samples are expected to correspond to the boundary between classes. This identification of boundary samples allows for the reduction of the learning set to a small relevant set (consisting of boundary samples), since these can be considered as the most representative samples for improving the classifier.

In the first iteration of the learning phase, the roots of the clusters are displayed to the user, who annotates their labels. These samples constitute the training set for the first instance of the classifier. For all other iterations, among the samples of the reduced set (boundary samples of the clusters) a few randomly chosen ones are selected for classification. Once classified, these samples are submitted to the user for confirmation of the labels assigned by the current classifier. Since only a small number of misclassified samples require annotation, the user's time and effort are lessened. In fact, as the classifier improves throughout the iterations the actions required from the user are increasingly reduced. After the labels are confirmed/corrected by the user, the samples are incorporated into the training set and a new instance of the classifier is generated. This entire cycle is repeated until the user is pleased with the accuracy of the classification.

Moreover, it is important to emphasize that different clustering techniques (such as k-means or k-medoids) can be used in the data reduction phase. Similarly, different supervised classifiers can be used in the classification and selection phases of the proposed paradigm. We choose OPF-based clustering since it offers many advantages, as mentioned in Section 2.

5 EXPERIMENTS

For evaluation, we developed a baseline approach (OPF-Rand) using the OPF-classifier and random selection of samples. At each learning iteration, the same number of random samples is selected from the entire dataset for OPF-Rand and, from the reduced dataset, for the Cluster-OPF-Rand. This number of samples is equal to the number suggested by Cluster-OPF-Rand based on the clustering results – a fair choice. These samples are classified and presented to the user for annotation. The user annotates the misclassified samples and they are added to the training set to improve the OPF classifiers used in each method for the next iteration. Thus, one can easily note the gain obtained by using clustering for dataset reduction, which induces the knowledge of a large number of classes, resulting in an early increase in accuracy. Moreover, clustering also allows for the choice of random samples from the reduced set comprised of good representative samples, instead of a much larger set of data (as in OPF-Rand).

The reported results were compiled from the average of experiments run 10 times, with randomly generated learning sets and unseen test sets for accuracy measures. For all datasets used, we chose 80% of the available samples for learning, and 20% for testing.

5.1 The Dataset Description

To perform the experiments we have used real-world datasets from very diverse domains. Due to space limitations, in the present paper there are only results obtained from three datasets.

The first dataset was obtained from the University of Notre Dame [8]. It was originally designed to study the effect of time on face recognition. The images were acquired in several weekly sessions with the participation of distinct individuals. In these sessions, different expressions (neutral, smiling, sad) were captured. In this work, we concentrated on a subset containing 1,864 samples with 162 features and 54 classes. Figure 4 displays specimens from this dataset.

Figure 4: Examples of images from the Faces dataset.

The second dataset is composed of images of parasites, provided by a research laboratory at the University of Campinas, where faecal parasitological examination is performed for diagnosis of enteroparasitosis present in humans. We used a dataset consisting of 1,660 faecal samples with 262 features and 15 classes. A particularity of this set is that each class contains a different number of images varying from 33 to 163 depending on the parasite species found on microscope slides. Figure 5 displays samples from this dataset.

The third one is the Pen-Based Recognition of Handwritten Digits dataset obtained from the UCI Machine

Faces	Accuracy (%)		Total Annotated Images (%)	
Iteration	Cluster-OPF-Rand	OPF-Rand	Cluster-OPF-Rand	OPF-Rand
1	94.85	85.11	6.51	6.51
2	97.27	94.21	7.59	8.51
3	98.06	97.35	8.11	9.40
4	98.57	98.35	8.41	9.78
5	98.85	98.78	8.68	9.98

Table 1: Accuracies and total annotated images for Cluster-OPF-Rand and OPF-Rand on the Faces dataset.

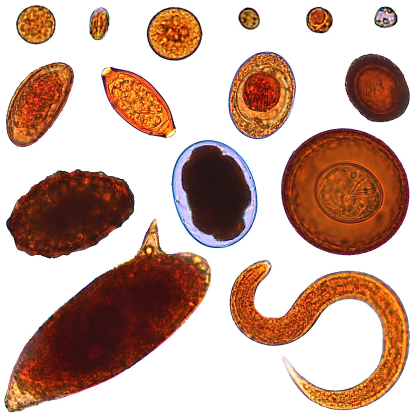


Figure 5: Examples of images from each class of the structures of intestinal parasites in the Parasites dataset.

Learning Repository [15], that consists of 10,992 objects in 16 dimensions, distributed in 10 classes corresponding to the digits [0...9]. The 16 dimensions are drawn by re-sampling from handwritten digits. This digits database was built from a collection of 250 samples from 44 writers.

5.2 Results

To compare the effectiveness of each method (Cluster-OPF-Rand and OPF-Rand), Tables 1-3 present the mean accuracy and the total annotated images using the datasets Faces, Parasites and Pendigits, respectively. It is important to emphasize that comparisons were not performed between Cluster-OPF-Rand and methods that require classifying and organizing *all* samples in the database, at each learning iteration, due to this process being infeasible in practice.

Notice that the proposed method creates a new classifier instance at each iteration. We would like to verify the ability of Cluster-OPF-Rand in choosing the most representative samples from a reduced set, as well as, in which iteration, whether the user might be pleased with the classification accuracy. Therefore, we monitor the mean accuracy of each instance on the unseen samples of the test set. Furthermore, for each sample set selected at each iteration, we simulate the user interaction by correcting the misclassified labels given by the current classifier instance. Tables 1-3 help compare the total number of annotated images used to increase the training set.

In summary, Cluster-OPF-Rand started off with a better performance than OPF-Rand, for all datasets analysed. Moreover, it achieves high accuracies sooner. To reach the same accuracies, the randomized method (OPF-Rand) required more samples annotated by the user as well as more learning iterations than Cluster-OPF-Rand.

Using the Faces dataset (Table 1), both methods achieve similar accuracies and both can be improved with more user annotations and more learning iterations. However, Cluster-OPF-Rand allows the learning process to stop earlier in comparison with OPF-Rand. Furthermore, it is important to highlight that, out of 1,469 samples only 132.94 (about 9.05%) had to be annotated for the proposed method to achieve accuracy above 99%, in its last (9^{th}) iteration using all samples on the reduced set. These results are similar to those for the remaining datasets (Tables 2 and 3). This shows that our method can outperform OPF-Rand in effectiveness.

Considering the Parasites dataset (Table 2), in the first iteration, Cluster-OPF-Rand achieves accuracies above 92% with less than 2% of the learning samples annotated by the user, while the randomized method OPF-Rand reaches similar accuracies only from the fourth iteration on and requiring the user to annotate more than 3% of the learning samples. Furthermore, out of 1,323 samples only 77.7 (about 5.87%) had to be annotated for Cluster-OPF-Rand to achieve an accuracy above 97%, in its last (25^{th}) iteration using all samples in the reduced set.

For the Pendigits dataset (Table 3), our method obtains high accuracies in all learning iterations. In the first one, it presents an accuracy of 88.80%. In the remaining iterations, the accuracies tend to increase continuously, reaching over 99%. Furthermore, out of 8,791 samples only 79.9 (about 0.90%) had to be annotated for the proposed method to achieve accuracy above 97% in the 30^{th} iteration. In a practical situation, a user would be very pleased at this point, mainly considering that the randomized method (OPF-Rand) learning process consists of 440 iterations, when using all available learning samples.

Figure 6a-b illustrates the mean accuracies and the number of samples annotated by the user at each iteration for each dataset using Cluster-OPF-Rand,

Parasites	Accuracy (%)		Total Annotated Images (%)	
Iteration	Cluster-OPF-Rand	OPF-Rand	Cluster-OPF-Rand	OPF-Rand
1	92.68	79.44	1.98	1.98
2	94.12	88.50	2.54	2.66
3	94.94	91.60	2.91	3.06
4	95.30	92.67	3.12	3.29
5	95.21	93.64	3.36	3.54

Table 2: Accuracies and total annotated images for Cluster-OPF-Rand and OPF-Rand on the Parasites dataset.

Pendigits	Accuracy (%)		Total Annotated Images (%)	
Iteration	Cluster-OPF-Rand	OPF-Rand	Cluster-OPF-Rand	OPF-Rand
1	88.80	70.36	0.13	0.13
2	90.96	82.97	0.22	0.25
3	91.99	87.49	0.29	0.30
4	92.89	89.72	0.35	0.35
5	93.70	91.25	0.40	0.40

Table 3: Accuracies and total annotated images for Cluster-OPF-Rand and OPF-Rand on the Pendigits dataset.

respectively. We used logarithmic scales, due to the size of these datasets. Our method requires a greater effort by the user in the first few iterations, since the selected samples are the most difficult to classify. However, looking at the end of the learning phase, one can observe that the proposed method demands less effort from the user, who annotates much fewer samples after some iterations (reaching almost no annotations at all).

The reduction strategy becomes very important in a process where a goal is to limit the number of iterations to as few as possible. In this context, selecting samples that speed up the improvement of the classifier through the iterations becomes critical. The more difficult to classify the selected samples in the current iteration are, the more useful they are to improve the classifier for the next iteration. Therefore, the selection of hard to classify samples coupled with the early knowledge of all classes allow for higher accuracy sooner.

Note that, in the first iteration with all datasets (Tables 1-3), Cluster-OPF-Rand provides higher accuracies than OPF-Rand. Using roots of each cluster for the first classifier instance becomes really important due to its use in the next iteration. This reduces the time and effort by the user who mainly has only to confirm the labels of the samples that have already been classified. Hence, this first instance of the classifier should be based on the knowledge of as many classes as possible (ideally, all of them). In later learning iterations, the performance gain depends on the choice of good samples. With the proposed method, it is possible to improve these choices by reducing a large dataset to a small subset consisting of boundary cluster samples for the training of the subsequent classifiers.

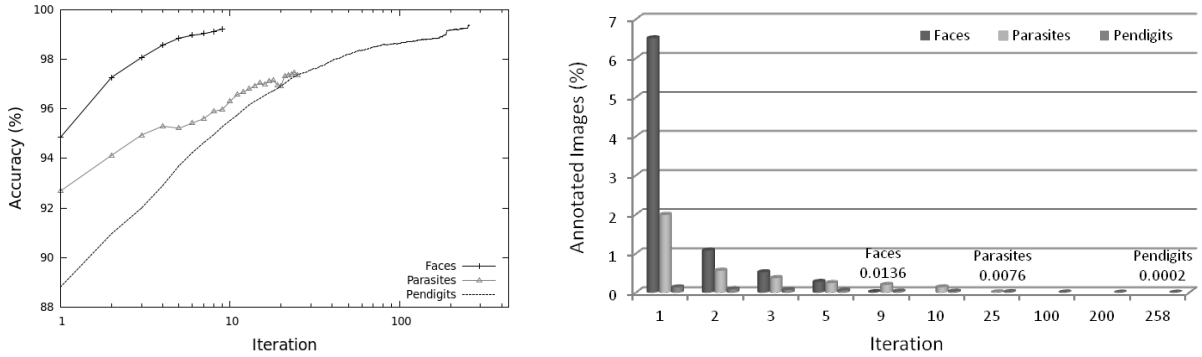
It is clear that Cluster-OPF-Rand, in addition to providing high accuracies, requires fewer learning iterations than those demanded by OPF-Rand. Additionally, it relies on fewer interactions with the user whose effort is reduced to almost none after a few iterations. Therefore, clustering improves the knowledge of samples from most/all classes. From the results presented, we can see that clustering roots allow us to obtain high accuracy since the first iteration. In the remaining iterations, the growth of accuracy is faster for Cluster-OPF-Rand, which also proves beneficial for the reduction strategy proposed.

6 CONCLUSION AND FUTURE WORK

In this work, we introduced an efficient active learning paradigm which enables the reduction and organization of the learning set a priori. A first instantiation, Cluster-OPF-Rand, of the proposed paradigm was developed in order to illustrate its effectiveness. The data reduction is based on clustering and the organization uses a randomized choice of samples of the reduced set, which contains the most representative (boundary) ones for the learning process. Cluster-OPF-Rand enables us to achieve the desired results, by using the knowledge of both user and classifier, at each learning iteration, along with the reduction strategy developed.

We concluded that our paradigm is more suitable to handle large datasets than the traditional one where methods require, at each learning iteration, the classification of *all* samples in the database, followed by their organization, and, finally selection. The proposed paradigm enables the reduction and organization phases to occur only once, as pre-processing. In addition, classification does not occur for all samples in the database, but to a small set of samples.

Experiments with datasets from distinct applications showed that Cluster-OPF-Rand, in addition to achiev-



(a) (b)

Figure 6: Comparison of Cluster-OPF-Rand on the three datasets. (a) Mean accuracy on the test sets. (b) Total annotated samples in each iteration (in percentage).

ing higher accuracy sooner, requires fewer learning iterations than those presented by OPF-Rand. Moreover, it is important to highlight that the user's time and effort are reduced to almost none after just a few iterations. Furthermore, experiments also demonstrated that it is possible to reduce the user's effort by over 90%, obtaining a classification accuracy above 97%.

Considering that new technologies have provided large datasets for many applications and that he traditional paradigms for active learning present unacceptable training times, we conclude that the proposed paradigm is an important contribution to active machine learning. Future works include developing other ways to explore the reduction and organization of data, for instance, a strategy that relies on an absolute criterion established a priori which explores the organization of the data in the feature space.

7 ACKNOWLEDGEMENTS

This work has been supported by grants from Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq): 481556/2009-5, 303673/2010-9, 552559/2010-5, 483177/2009-1, 473867/2010-9; from Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES): 01-P-01965/2012; from Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP): 07/52015-0 and from FAEPEX/UNICAMP.

8 REFERENCES

[1] D. Angluin. Queries and Concept Learning. *Machine Learning*, 2:319–342, 1988.

[2] F.A.M. Cappabianco, J.S. Ide, A.X. Falcão, and C.-S.R. Li. Automatic subcortical tissue segmentation of mr images using optimum-path forest clustering. In *International Conference on Image Processing (ICIP)*, pages 2653–2656, 2011.

[3] Yizong Cheng. Mean shift, mode seeking, and clustering. *TPAMI*, 17(8):790–799, 1995.

[4] D. A. Cohn, Z. Ghahramani, and M. I. Jordan. Active learning with statistical models. *JAIR*, 4:129–145, 1996.

[5] A. T. da Silva, A. X. Falcão, and L. P. Magalhães. A new CBIR approach based on relevance feedback and optimum-path forest classification. *Journal of WSCG*, pages 73–80, 2010.

[6] A. T. da Silva, A. X. Falcão, and L. P. Magalhães. Active learning paradigms for CBIR systems based on optimum-path forest classification. *Pattern Recognition*, 44:2971–2978, 2011.

[7] D. T. Davis and J. N. Hwang. Attentional focus training by boundary region data selection. In *Intern. Joint Conference on Neural Networks (IJCNN)*, volume 1, pages 676–681, 1992.

[8] Faces. Biometrics Database Distribution. The Computer Vision Laboratory, University of Notre Dame, 2011. `www.nd.edu/~cvrl/CVRL/Data_Sets.html`.

[9] A. Holub, P. Perona, and M.C. Burl. Entropy-based active learning for object recognition. In *CVPRW*, pages 1–8, 2008.

[10] P. Jain and A. Kapoor. Active learning for large multi-class problems. In *IEEE Conference on Computer Vision and Pattern Recognition(CVPR)*, pages 762–769, 2009.

[11] A. Kapoor, K. Grauman, R. Urtasun, and T. Darrell. Gaussian Processes for Object Categorization. *International Journal of Computer Vision (IJCV)*, 88:169–188, 2010.

[12] X. Li, L. Wang, and E. Sung. Multi-label SVM Active Learning for Image Classification. In *International Conference on Image Processing (ICIP)*, volume 4, pages 2207–2210, 2004.

[13] J. P. Papa, A. X. Falcão, V. H.C. de Albuquerque, and J. M.R.S. Tavares. Efficient supervised optimum-path forest classification for large datasets. *Pattern Recognition*, 45:512–520, 2012.

[14] J. P. Papa, A. X. Falcão, and C. T. N. Suzuki. Supervised pattern classification based on optimum-path forest. *Intern. Journal of Imaging Systems and Technology (IJIST)*, 19(2):120–131, 2009.

[15] Pendigits. Pen-Based Recognition of Handwritten Digits Dataset. UCI - Machine Learning Repository, 2011. `archive.ics.uci.edu/ml/datasets/Pen-Based+Recognition+of+Handwritten+Digits`.

[16] G.-J. Qi, X.-S. Hua, Y. Rui, J. Tang, and H.-J. Zhang. Two-dimensional multilabel active learning with an efficient online adaptation model for image classification. *IEEE Transact. on Pattern Analysis and Machine Intel.*, 31(10):1880–1897, 2009.

[17] L. M. Rocha, F. A. M. Cappabianco, and A. X. Falcão. Data clustering as an optimum-path forest problem with applications in image analysis. *Intern. Journal of Imaging Systems and Technology (IJIST)*, 19(2):50–68, 2009.

[18] S. Tong and E. Chang. Support vector machine active learning for image retrieval. In *ICM*, pages 107–118. ACM, 2001.

[19] S. Tong and D. Koller. Support vector machine active learning with applications to text classification. *Journal of Machine Learning Research (JMLR)*, 2:45–66, 2002.

[20] R. Yan, J. Yang, and A. Hauptmann. Automatically labeling video data using multi-class active learning. In *IEEE Intern. Conference on Computer Vision (ICCV)*, volume 1, pages 516–523, 2003.