CAN WE AUTOMATICALLY CHOOSE BEST UNCERTAINTY HEURISTICS FOR LARGE MARGIN ACTIVE LEARNING?

Ines BEN SLIMENE BEN AMOR (a,c), Nesrine CHEHATA (b), Philippe LAGACHERIE (c), Jean-Stéphane BAILLY (d) and Imed Riadh FARAH (a)

ines.ben-slimane@supagro.inra.fr - Nesrine.Chehata@ensegid.fr - philippe.lagacherie@supagro.inra.fr - jean-stephane.bailly@agroparistech.fr - riadh.farah@ensi.rnu.tn

(a) RIADI Laboratory, ENSI, Manouba, Tunisia
(b) Bordeaux INP, G&E,EA 4592, France
(c) INRA, UMR LISAH, 1221 Montpellier, France
(d) AgroParisTech, UMR LISAH, 1221 Montpellier, France

ABSTRACT

Active learning (AL) has shown a great potential in the field of remote sensing to improve the efficiency of the classification process while keeping a limited training dataset. Active learning uses heuristics to select the most informative pixels in each iteration. In literature, there are several metrics and selection criteria. In this paper, we focus on the uncertainty heuristics for large margin active learning. Existing uncertainty metrics are presented and combined to new ones using support vector machine learning algorithm. Besides, a new methodology is proposed, which automates a priori the choice of the best uncertainty heuristic. This contribution is evaluated on hyperspectral datasets while varying two parameters: class mixing and class balance. Finally discussion and conclusion are drawn.

Index Terms— Active learning, large margin, uncertainty, support vector machine, metrics.

1. INTRODUCTION

Statistical learning models are intensively used by remote sensing community [1]. Support Vector Machines (SVMs), neural networks and Random Forest are the most considered for the land use classification. However, supervised algorithms performance depends on data representation used to build the models classifier. This constraint makes generating an efficient learning set more difficult and expensive, requiring a deep manual image analysis or field surveys and successive labelling of each pixel [2].

Defining an efficient learning set is a key issue in remote sensing image classification. Limited financial and temporal resources, the complexity of algorithms and a high intraclass variance performed with a sub-optimal dataset, can make an algorithm fail. Active learning aims to build an efficient learning set by improving the model performance by iteratively growing learning set.

The classification model is adapted regularly by adding new labelled pixels which are most beneficial for improving the model performance. The active learning strategy has been, at a vast majority, used with SVM classification [3]. To identify the most informative pixels, it needs a strategy to rank candidate pixels. Two criteria are often coupled: uncertainty and diversity. The uncertainty criterion is related to the algorithm confidence in correctly classifying a pixel. The diversity criterion ensures that the learning pixels are different from each other. The existing methods can be grouped into three main families [2]: (1) query by-committee ; (2) posterior probability and (3) large margin heuristics. In [4], the three main AL families are presented with a comparative study focused on uncertainty methods and based on [2]. This study is focused on the large margin active learning and especially the uncertainty heuristics.

In literature, the key issue is the fact that different uncertainty measures are used and behave differently with regard to various hyperspectral data. The study tries to automate the choice of the best uncertainty measures based on the first iteration. The methodology is evaluated on hyperspectral data with two varying parameters: class mixing and class balance.

2. UNCERTAINTY HEURISTICS OF LARGE MARGIN AL

This family is based on support vector machine (SVM) method. The separation distance of the hyperplane is a simple way to estimate the model confidence on an unlabelled pixel. A minimum distance means that the pixel is close to the hyperplane, giving a maximal uncertainty. In a binary classification, the distance between a pixel and the SVM hyperplane is given by:

$$f(x_i) = \sum_{j=1}^{n} \alpha_j y_j K(x_j, x_i) + b \qquad (1)$$

Where $K(x_j, x_i)$ is a kernel, which defines the similarity between the candidate x_i and the Support Vectors (SV) x_j, which are the pixels showing non zero α_j coefficients. And y_j is the labels of the support vectors. The heuristic that takes advantage of this property is called Margin Sampling (MS). Three heuristics illustrate the corresponding uncertainty criterion; MS, MCLU and SSC. They are summarized in table 1.

2.1. Margin Sampling (MS)

The MS heuristic takes benefits of the SVM geometric properties by the fact that SVs are labeled examples which are related to the margin with a value equal to one. MS performs a candidate sampling that minimizes $f(x_i)$ [5].

2.2. MultiClass Level Uncertainty (MCLU)

The MCLU is extended from MS to solve the uncertainty problem of multi classes. There are two functions to minimize [6] :

- The function $f(x_i, w)$ which is the pixel distance relative to the hyperplane and defined for the class w in the case of multiple classes.

- The function $f(x_i)^{MC}$ which gives a confidence value. Instead of considering the most uncertain SVM class, MCLU estimates the difference between the distances of the two most likely classes relative to the margin.

2.3. Significance Space Construction (SSC)

The uncertainty in this heuristic is measured by using a coefficient in order to convert the multi-class classification problem into a binary one. This results in the definition of a new classification function that allows to choose pixels which are most likely to become SV [2].

AL	R	H	Advantage	Disadvantage
Large margin	[5]	MS	More efficient if initial training set is small.	-
	[6]	MCLU	Simpler and more efficient if mixed classes	-
	[7]	SSC	Simple and easy.	takes into account only SVs.

Table 1: Considered large margin uncertainty heuristics (**R=** Reference, **H=** Heuristic).

3. METHODOLOGY

In practice, uncertainty metrics performances vary highly with the dataset. To tackle this issue, this work proposes two novelties: 1) the existing MS and MCLU heuristics are combined and two new metrics are proposed, 2) an automation of the best uncertainty heuristic choice since the first AL iteration is proposed.

3.1. Proposed uncertainty heuristics

Two large margin uncertainty heuristics are proposed which combine MS and MCLU metrics. All metrics are normalized by their maximum. At each iteration, candidate pixels are sorted with regard to the considered metric, a batch of n most uncertain pixels is then selected and added to the training set.

MEAN: This metric is designed to select pixels that have a low average value between MS and MCLU results assuming that uncertain pixels will have low values in both metrics.

MIN: This heuristic aims to choose pixels that have the lowest uncertainty value between MS and MCLU heuristics. Then, the lowest values of MIN metric lead to the most uncertain pixels.

3.2. Datasets

In order to evaluate our methodology, the idea is to use an existing original hyperspectral dataset and to modify its parameters to simulate new datasets. Two parameters are considered: class mixing and class balance. Among the literature hyperspectral datasets, Pavia university one was chosen since it is unbalanced and has well distinguished classes which allows us to simulate two other datasets.

PaviaU : It has been acquired by ROSIS sensor at a 4 nm spectral resolution and a 1.3 m spatial resolution. The image consists of 103 spectral bands. The ground truth differentiates 9 classes.

PaviaU-MUB: Mixed and UnBalanced classes. Pixels belonging to similar thematic classes are grouped leading to mixed classes.

PaviaU-UMB: UnMixed and Balanced classes. The same number of pixels will be kept for each class.

3.3. Margin histogram indicators

Our assumption is to say that the metric whose distribution has more uncertain pixels will be more efficient for the AL process. The analysis of each metric distribution at the first AL iteration is processed using different histogram indicators that will helps us choosing the best metric. The tested indicators are:

- **Average** (A): The mean value of the distribution.

- **Standard Deviation** (SD) : measures the amount of variation or dispersion from the average.

- **Skewness (S)**: is an indicator of distribution asymmetry. If $S > 0$: Right skewed distribution where most values are concentrated on the left of the mean. If $S < 0$: Left skewed distribution, most values are concentrated on the right of the mean.

- **Kurtosis (K)**: is a measure of whether the data are peaked or flat relative to a normal distribution. Indeed, data sets with high kurtosis tend to have a distinct peak near the mean, while distribution with a flat top near the mean will lead to a low kurtosis.

- **First Quartile ($Q1$)** : The quartiles of a ranked set of data values are the three points that divide the data set into four equal groups. The first quartile (Q1) corresponds to the first quarter of the data.

4. EXPERIMENTS

4.1. Experimental Setup

In order to evaluate the proposed methodology, experiments were conducted on three pavia data sets, described above. For each experiment, the reference pixels were split in three sets, corresponding to initial balanced training set L_0 with $Card(L_0) = 45$, the unlabeled candidate set U with $Card(U) = 700$, and the test set. The AL algorithm is applied with SVM algorithm.

In this paper, we consider a photo-interpretation context (as often in literature), thus the number of iterations nIt can be important and the number of added pixels at each iteration (Batch B) can be small. In our experiments, we fixed $nIt = 20$ and $Card(B) = 5$.

4.2. Results

All uncertainty heuristics; MS, MCLU, MEAN and MIN are applied for the first AL iteration on the three datasets. Results are presented from the easiest to the more complicated cases.

4.2.1. *PaviaU-UMB*

This Pavia simulation has unmixed classes and a balanced dataset. Table 2 summarizes the histogram indicators for all metrics. Figure 1(a) illustrates the classification Kappa values with AL iterations for all metrics.

When analysing both results, one can observe that most pertinent indicators are Average (A) and Quartile $Q1$. The lower they are, the more uncertain pixels are. Metrics with the lowest A et $Q1$ values will be considered as the most efficient for AL process on the current data. In this case, it appears to be the MIN metric, which is also confirmed by the Kappa values. On the contrary, the MS metric with the highest A and $Q1$ values leads to the lowest kappa values. However, in this easy case, all metrics tend to have similar Kappa values while MIN metric seems to be the most efficient at the end of the process. Moreover, for all metrics, the AL process allows to increase the kappa values by almost 10%.

Heuristic	A	SD	S	K	Q1
MS	0,63	0,32	-0,44	1,74	0,33
MCLU	0,51	0,25	-0,46	2,01	0,30
MEAN	0,57	0,28	-0,42	1,69	0,29
MIN	**0,44**	0,24	-0,33	1,93	**0,23**

Table 2: Distribution indicators for PaviaU-UMB simulation

4.2.2. *PaviaU*

For the original data with unmixed classes and a unbalanced dataset. Distribution indicators are summarized in table 3. AL classification performances are illustrated on figure 1(b).

Once again the MIN metric leads to the lowest A and $Q1$ values, showing more uncertain pixels. The performance MIN metric is confirmed by Kappa values since it is the most efficient metric both at the beginning and end of AL process. Finally, the contribution of AL process is higher in more complicated cases. The kappa value is increased by almost 25%.

Heuristic	A	SD	S	K	Q1
MS	0,48	0,31	0,02	1,61	0,19
MCLU	0,40	0,23	0,01	1,97	0,19
MEAN	0,49	0,29	-0,01	1,58	0,20
MIN	**0,38**	0,24	0,01	1,81	**0,14**

Table 3: Distribution indicators for PaviaU dataset.

4.2.3. *PaviaU-MUB*

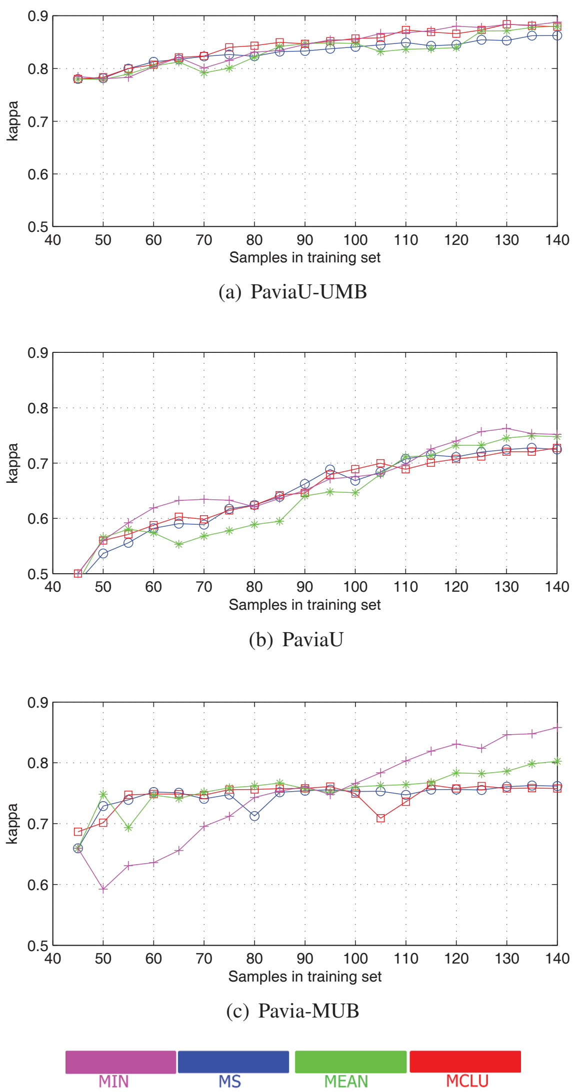
This unbalanced dataset with mixed classes is the more complicated study case. The distribution indicators are given in table 4. AL classification performances are reported on figure 1(c).

From the indicator table, one can see that the MIN metric has again the lowest values for both A and $Q1$ and its higher performance is confirmed by the Kappa values.

For this worst study ca,s the kappa reaches 0.8 improving the initial result by 20%.

Heuristic	A	SD	S	K	Q1
MS	0,57	0,31	-0,23	1,63	0,29
MCLU	0,44	0,23	0,05	2,29	0,28
MEAN	0,47	0,27	0,10	1,84	0,23
MIN	**0,43**	0,27	0,02	1,84	**0,19**

Table 4: Distribution indicators for PaviaU-MUB simulation



(a) PaviaU-UMB

(b) PaviaU

(c) Pavia-MUB

Fig. 1: AL Classification performance of uncertainty metrics. (a) PaviaU-UMB dataset, (b) PaviaU dataset and (c) PaviaU-MUB dataset

These experiments allow to conclude on three points: 1) the efficiency of the proposed MIN metric that combines MS and MCLU ones 2) the automatic selection of uncertainty metric is possible using metric distribution indicator 3) the contribution of AL process is more important in complex classification problem.

5. CONCLUSION

In this paper, new AL uncertainty heuristics, that are a combination of MS and MCLU ones were proposed. Keeping pixels with the minimal value for both metrics led to the best classification performances. Results were validated on derived hyperspectral data from PaviaU dataset varying class mix and class balances parameters. In addition, we proposed to automate the choice of the uncertainty metric, which is data-dependent. Our proposed method is based on the distribution indicators of the 1st AL iteration. Average and first quartile seems to be the most efficient indicators to predict the best uncertainty heuristics. This strategy offers the advantageous automated prediction of the most efficient metrics only from the first AL iteration.But, the proposed strategy is only valid with the uncertainty criteria. Similar work has to be done using diversity criteria. Moreover, as future work, more simulation will be processed using other parameters and on different datasets.

6. REFERENCES

[1] Trevor Hastie, Robert Tibshirani, Jerome Friedman, and James Franklin, "The elements of statistical learning: data mining, inference and prediction," *The Mathematical Intelligencer*, vol. 27, no. 2, pp. 83–85, 2005.

[2] D. Tuia, M. Volpi, L. Copa, M. Kanevski, and J. Munoz-Mari, "A survey of active learning algorithms for supervised remote sensing image classification," *IEEE Journal of Selected Topics in Signal Processing*, vol. 5, no. 3, pp. 606–617, June 2011.

[3] S. Patra and L. Bruzzone, "A batch-mode active learning technique based on multiple uncertainty for SVM classifier," *IEEE GRSL*, vol. 9, no. 3, pp. 497–501, 2012.

[4] I. Ben Slimene, N. Chehata, I.R. Farah, and P. Lagacherie, "Uncertainty heuristics of large margin active learning for hyperspectral image classification," in *Image Processing, Applications and Systems Conference (IPAS), 2014 First International*, Nov 2014, pp. 1–6.

[5] Pabitra Mitra, B. Uma Shankar, and Sankar K. Pal, "Segmentation of multispectral remote sensing images using active support vector machines," *Pattern Recognition Letters*, vol. 25, no. 9, pp. 1067 – 1074, 2004.

[6] B. Demir, C. Persello, and L. Bruzzone, "Batch-mode active-learning methods for the interactive classification of remote sensing images," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 49, no. 3, pp. 1014–1031, March 2011.

[7] E. Pasolli and F. Melgani, "Model-based active learning for svm classification of remote sensing images," in *Geoscience and Remote Sensing Symposium (IGARSS), 2010 IEEE International*, July 2010, pp. 820–823.