Probabilistic Breaking Tie: An Active Learning Strategy To Leverage Class Hierarchy For Impervious Surfaces Classification

Romain Thoreau, Véronique Achard, Laurent Risser, Béatrice Berthelot, Xavier Briottet

PROBABILISTIC BREAKING TIE: AN ACTIVE LEARNING STRATEGY TO LEVERAGE CLASS HIERARCHY FOR IMPERVIOUS SURFACES CLASSIFICATION

Romain Thoreau[*,1,2], *Véronique Achard*[1], *Laurent Risser*[3], *Béatrice Berthelot*[2] *and Xavier Briottet*[1]

[1] ONERA, The French Aerospace Lab, 2 Av. Edouard Belin, 31055 Toulouse, France
[2] Magellium, 1 rue Ariane, 31520 Ramonville Saint-Agne, France
[3] Toulouse Mathematics Institute (UMR 5219), CNRS, University of Toulouse, 31062 Toulouse, France

ABSTRACT

Urban sprawl has significantly changed the land cover of metropolis and periurban areas. In that regard, the monitoring of impervious surfaces has become crucial to mitigate their impacts on the hydrology of the watersheds and on urban micro-climates. At this end, airborne hyperspectral imaging is very well suited to map the diversity of impervious surfaces over urban areas, both in terms of scale and of discriminative power. Yet, the diversity of soil materials as well as the similarity between some artificial porous and impervious classes, is such that active learning methods are imperative to build optimal ground truths for machine learning algorithms. Besides, in the context of impervious surfaces classification, not all classes are of equal interest. In the present paper, we introduce Probabilistic Breaking Tie, a query system, built on the state-of-the-art Breaking Tie [1] heuristic, that leverages class hierarchy as *a priori* semantic knowledge. Our numerical experiments on the Houston University data set [2] show that our method significantly increases the pace at which active learning improves the accuracy metrics over impervious and porous classes. Code is available at https://github.com/Romain3Ch216/AL4EO.

Index Terms— Hyperspectral imaging, semantic segmentation, impervious surfaces, active learning.

1. INTRODUCTION

In past decades, the urban sprawl has significantly changed the land cover of metropolis and periurban areas [3]. As most artificial surfaces have low permeability coefficients and low albedo, the hydrology of the watersheds as well as urban micro-climates have undergone major changes [4, 5]. Machine learning algorithms have been rather successful in producing imperviousness maps from satellite images, such as the *Imperviousness Density* product from the European Copernicus program derived from Sentinel-2 data. However, the poor spatial and spectral resolutions of Sentinel-2 does not enable the distinction of different types of impervious surfaces. Scientific studies and urban policies could, although, benefit from a fine nomenclature of porous and impervious surfaces. For instance, discriminating impervious materials based on their permeability coefficient [6], on their ageing [7] or even on their connection to the sewer system [8], as well as distinguishing porous classes such as trees, grass or bare soils could be useful for urban management [9].

Airborne hyperspectral imaging is very well suited to distinguish such a diversity of materials due to its high spectral and spatial resolutions. However, building training data sets with such a fine nomenclature is very challenging and requires expensive field campaigns. Consequently, only a limited number of pixels can be labeled, hence the need to choose the most informative ones. This is the goal of Active Learning (AL) methods: an iterative algorithm selects the *best* pixels to be labeled [10]. The limitation of AL methods however is that they give equal importance to each class. In the context of impervious surfaces classification though, we would rather well recognize impervious roads from porous railways rather than healthy grass from stressed grass. Therefore, we would like to avoid the *waste* of annotations on pixels of little importance.

In this paper, we leverage the natural hierarchical structure of porous and impervious classes to perform Active Learning. Our method, Probabilistic Breaking Tie, builds on the state-of-the-art Breaking Tie [1] heuristic to prioritize pixels that are informative of the differences between porous and impervious classes.

Section 2 introduces Probabilistic Breaking Tie in detail. Section 3 presents the numerical experiments and the results. We finally conclude in section 4.

2. METHOD

2.1. The Active Learning Framework

Active Learning (AL) algorithms depend on an acquisition function (also named the query system), an oracle (a user that labels the data), a number of iterations and a budget of pixels to query per iteration. In the following, we denote by L_t

* Corresponding author: romain.thoreau@onera.fr

and U_t the labeled and unlabeled sets respectively at step t. The data points $x_i \in \mathcal{X}$ for all $i \in \{1, \ldots, N\}$ are spectral or spectral-spatial features, where N is the total number of pixels. The acquisition function a usually depends on L_t and on a classifier f that takes an input x and outputs a probability distribution over C classes. At first, a small training data set is manually labeled. Then, at each iteration t, a model f is trained from scratch, and a subset of the unlabeled pool U_t is selected in order to maximize, or minimize, the acquisition function. Newly labeled pixels by the oracle are added to the training data set and the process repeats until the final step is reached.

2.2. Breaking Tie (BT) [1]

A classifier $f : \mathcal{X} \longrightarrow [0, 1]^C$ outputs a probability distribution over C classes from a data point $x \in \mathcal{X}$. The breaking tie heuristic for a data point x is a measure of confidence:

$$a_{BT}(x) = \max_{i \in (1,\ldots,C)} f(x)^i - \max_{j, j \neq i} f(x)^j \tag{1}$$

where $f(x)^i$ denotes the i^{th} coordinate of $f(x)$, *ie* the estimated probability that x belongs to class i. The function a_{BT} takes low values when a data point lies near a decision boundary.

2.3. Probabilistic Breaking Tie (PBT)

We introduce Probabilistic Breaking Tie (PBT), an enhancement of Breaking Tie to leverage class hierarchy.

The class hierarchy can be formalized by a tree (cf fig. 1), where the leaves are classes. Like in [11] and [12], we use the distance between leaves to encode the semantic similarity between classes. The lower the distance, the higher the similarity. Similarly, the higher the distance, the higher the cost of confusion. We can now define a cost matrix, for C classes, as follows: $\forall (k, l) \in \{1, \ldots, C\}^2, k \neq l$,

$$D[k, l] = 10^{-\frac{\max_{u,v} d(u,v) - d(k,l)}{\beta}} \tag{2}$$

where β is a penalty coefficient, $d(k, l)$ is the distance in the tree between the leaves k and l. $d(k, l)$ is the number of nodes within the path from leave k to the closest common parent with leave l. We can note that $d(k, l) = d(l, k)$ because every leaves are at the same depth in the tree.

This cost matrix is the core element of PBT. The idea is, instead of selecting b pixels with the lowest Breaking Tie scores, to sample b points according to a distribution defined by a weighted version of the Breaking Tie scores, where the weights are defined by the cost matrix.

Let's denote $X = \{x_i\}_{i \in (1,\ldots,N)}$ a pool of N points. We define the PBT acquisition function over a set of points as follows:

$$a_{PBT}(X) = \{\frac{w_i \cdot p_i}{\sum_{k=1}^{N} w_k \cdot p_k}\}_{i \in (1,\ldots,N)} \tag{3}$$

where, $\forall i \in (1, \ldots, N)$,

$$p_i = 1 - a_{BT}(x_i) \tag{4}$$

$$w_i = \delta_i D[k, l] \tag{5}$$

$$k,\ l = \underset{i \in (1,\ldots,c)}{\arg\max}\, f(x)^i,\ \underset{j, j \neq i}{\arg\max}\, f(x)^j \tag{6}$$

$$\delta_i = \begin{cases} 1 \text{ if } p_i > \gamma \\ 0 \text{ else} \end{cases} \tag{7}$$

with γ a threshold that controls the informativeness of the samples and D is the cost matrix.

Equation 4 defines p_i so that a low Breaking Tie score leads to a high probability to select pixel i. Equation 5 defines w_i so that it is lower when the two most likely classes have a low confusion cost. Equation 7 defines δ_i so that a pixel with a very high Breaking Tie score (*ie* not informative), cannot be selected despite the re-weighting.

3. NUMERICAL EXPERIMENTS

In the Active Learning comparative review [13], authors suggest to combine different AL strategies. They argue that representativeness heuristics should be use for the first iterations to have a good approximate of the data distribution and that inter-class uncertainty heuristic such as BT should be used at the last iterations in order to refine the decision boundaries between similar classes. Therefore, we are not interested in comparing PBT with representativeness or performance-based techniques but we experimentally study the benefits of PBT against BT when classes are organized hierarchically.

3.1. The Houston data set

The hyperspectral image over Houston university was acquired by the sensor ITRES CASI 1500 that covers the 380-1050 nm spectral range with 48 bands at a 1m ground sampling distance (GSD). More than four hundred thousands pixels were labeled over 20 classes (see fig. 1). However, we only kept in the initial training data set 200 labeled pixels per class, *i.e.* 4000 pixels in total, which is representative of an operational use case. Ten sets of different initial training data sets, unlabeled pools and test data sets were selected from disjoint regions.

We chose the Houston data set because it is, to our knowledge, the largest and the most complex labeled hyperspectral data set publicly available. In contrast to classic data sets such as Indian Pines or Pavia University, Houston has a rich hierarchical nomenclature. In future work though, we would like to test our method on other complex data sets. However, we

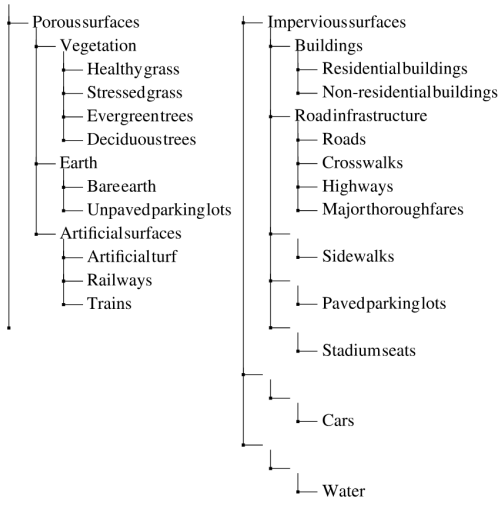

Fig. 1: Hierarchical nomenclature of the Houston data set

argue here that the ten splits of the ground truth yield very different initial training data sets, which should be enough to evaluate the robustness of PBT.

3.2. The model

To compute the BT and the PBT scores, we used a classic spectral CNN [14] that was benchmarked in the comparative review [15]. A better model would give better results, but we argue that the relative difference would be the same for BT and PBT. We performed 100 epochs with a 0.001 learning rate and a cross-entropy loss optimized through a stochastic gradient descent with momentum.

3.3. Metrics

At each step of the AL process (15 steps in total), we computed the following metrics based on the predictions of the CNN:

- **Overall Accuracy (OA)***. The number of correct predictions over the total number of predictions.
- **Mean Intersect Over Union (mIoU)***. The mean of the IoU score over every classes:

$$mIoU = \frac{1}{C}\sum_{k}^{C} IoU_k = \frac{1}{C}\sum_{k}^{C} \frac{TP_k}{TP_k + FP_k + FN_k} \tag{8}$$

  where TP_k, FP_k et FN_k respectively denote the true positives, the false negatives and the false positives for class k.

- **Confusion matrix**. We denote by Q^{true} and Q^{pred} the confusion matrices respectively normalized over the true labels and the predictions. In the following, we name and define the confusion matrix as:

$$Q = 0.5(Q^{true} + Q^{pred}) \tag{9}$$

- **Average Cost (AC)**. The average cost is the mean of the number of confusions weighted by their cost:

$$AC = \frac{1}{C^2}\frac{\sum_{i,j} Q[i,j] \cdot D[i,j]}{\sum_{i,j} D[i,j]} \tag{10}$$

  We normalize the average cost by the sum of the confusion cost matrix so that it is invariant to the costs scale.

* OA and mIoU were computed for N to N classification but also for porous versus impervious classification. In both cases, the model made predictions over C classes. Only then, the predicted labels were converted to *Porous*, *Impervious* or *Other*.

3.4. Hyperparameters

In our experiments, we set $\beta = 1$. This means that for two pixels whose breaking tie scores are equal, the pixel at the frontier of two classes closer from one node are ten times less likely to be queried. We set $\gamma = 0.8$ so that only uncertain pixels are queried.

3.5. Results

Fig. 2 shows that OA is barely unchanged between BT and PBT. On the contrary, mIoU was slightly increased by PBT, which was quite unexpected. It appears that PBT focused particularly on classes with high confusions rate.

Fig. 3 shows that PBT reached lower AC, especially at the beginning of the AL process. Similarly, impervious vs porous OA and mIoU were significantly increased by PBT. More specifically, PBT had a 4% higher mIoU at step 3. Moreover, PBT took two times less iterations than BT (and therefore two times less pixels to label) in order to reach a 0.64 IoU score.

Fig. 4a shows the difference of the confusion matrices obtained with PBT and BT at step 3. A zero value means that the proportion of confusions was unchanged between the two methods. A positive value on the diagonal means that more correct predictions were achieved with PBT. Negative values off-diagonal means that less confusions were achieved with PBT. For instance, we see that predictions of *Railways* are significantly increased with PBT by lowering the confusions with *Unpaved parking lots* (although they both belong to the *Porous* category). Finally, we note that the sum of the difference is nearly zero, which means that confusions and good predictions induced by PBT, compared to BT, balance out.

Fig. 4b shows the difference of the confusion matrices obtained with PBT and BT at step 3, weighted by the cost matrix. We can clearly see the benefits of PBT (for instance

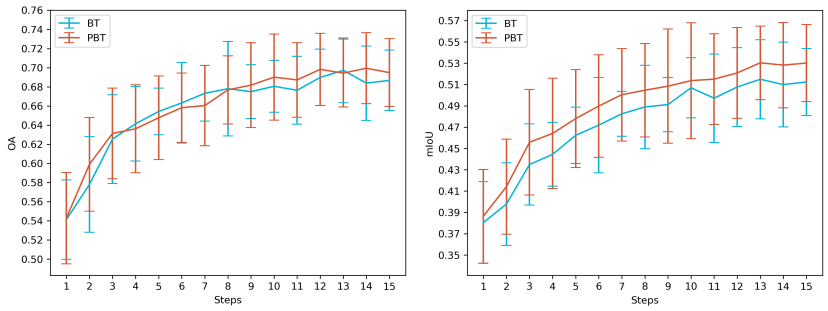

Fig. 2: Mean and standard deviation of OA and mIoU for N to N classification

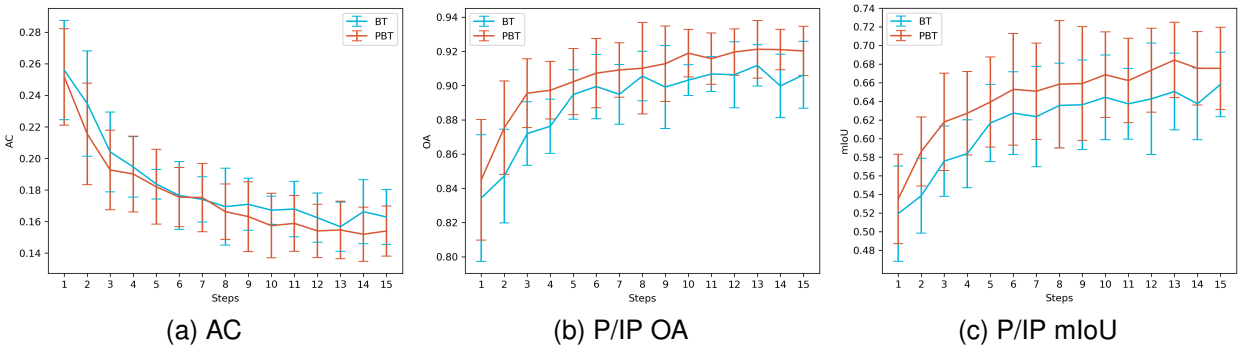

Fig. 3: Mean and standard deviation of AC, OA and mIoU for porous VS impervious classification

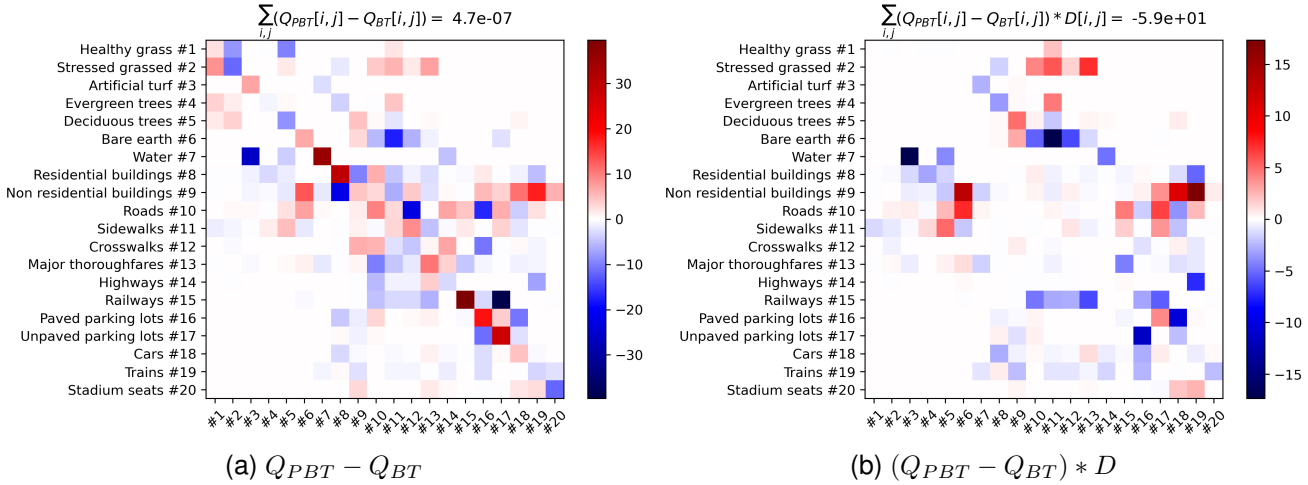

Fig. 4: (a): $Q_{PBT} - Q_{BT}$. (b): $(Q_{PBT} - Q_{BT}) * D$. Q_{PBT} and Q_{BT} denote the confusion matrices obtained with training data sets respectively labeled using the PBT and BT active learning methods after step 3. D denotes the cost of confusion matrix.

the decrease of confusions between *Bare earth* and *Sidewalks* but also its downsides (for instance the increase of confusions between *Non residential buildings* and *Trains*). Globally, the sum of the weighted difference is clearly negative, which demonstrates a decrease of the confusion cost.

4. CONCLUSION

We have introduced Probabilistic Breaking Tie, an Active Learning method that does not give the same importance to every classes, according to *a priori* semantic knowledge. This *a priori* knowledge was formalized by a hierarchical tree that results from the natural hierarchical structure of the land cover nomenclature. Numerical experiments on the complex hyperspectral Houston data set demonstrated the effectiveness of our method to decrease the average cost of confusion and to increase the accuracy metrics for impervious vs porous surfaces classification. In an operational context, it allows to reduce the number of pixels to label, hence the duration of field campaigns, in order to reach good performances. In future work, we will further study the impact of the stochasticity of Probabilistic Breaking Tie on the performance of the AL process.

5. REFERENCES

[1] Tong Luo, K. Kramer, S. Samson, A. Remsen, D. B. Goldgof, L. O. Hall, and T. Hopkins, "Active learning to recognize multiple types of plankton," in *Proceedings of the 17th International Conference on Pattern Recognition, 2004. ICPR 2004.*, 2004, vol. 3, pp. 478–481 Vol.3.

[2] Saurabh Prasad, Bertrand Le Saux, Naoto Yokoya, and Ronny Hansch, "2018 ieee grss data fusion challenge – fusion of multispectral lidar and hyperspectral data," 2020.

[3] Dennis Fox, Emmanuelle Witz, Violaine Blanc, Cécile Soulié, Marc Penalver-Navarro, and Alain Dervieux, "A case study of land cover change (1950–2003) and runoff in a mediterranean catchment," *Applied Geography*, vol. 32, pp. 810–821, 03 2012.

[4] Scott J. McGrane, "Impacts of urbanisation on hydrological and water quality dynamics, and urban water management: a review," *Hydrological Sciences Journal*, vol. 61, no. 13, pp. 2295–2311, 2016.

[5] V. Masson-Delmotte, P. Zhai, A. Pirani, S.L. Connors, C. Péan, S. Berger, N. Caud, Y. Chen, L. Goldfarb, M.I. Gomis, M. Huang, K. Leitzell, E. Lonnoy, J.B.R. Matthews, T.K. Maycock, T. Waterfield, O. Yelekçi, R. Yu, and B. Zhou (eds.), "Ipcc, 2021: Summary for policymakers. in: Climate change 2021: The physical science basis. contribution of working group i to the sixth assessment report of the intergovernmental panel on climate change," *Cambridge University Press. In Press.*, 2021.

[6] Tom Gleeson, Leslie Smith, Nils Moosdorf, Jens Hartmann, Hans Dürr, Andrew Manning, Ludovicus Beek, and Mark Jellinek, "Mapping permeability over the surface of the earth," *Geophysical Research Letters*, vol. 38, pp. L02401, 01 2011.

[7] Okan Bilge Özdemir, Hilal Soydan, Yasemin Yardımcı Çetin, and Hafize Şebnem Düzgün, "Neural network based pavement condition assessment with hyperspectral images," *Remote Sensing*, vol. 12, no. 23, pp. 3931, 2020.

[8] Maria del Rosario Viétez Vásquez, Birgitte Lilholt Sørensen, Ole Mark, Rasmus Lundgaard Borgstrøm, Kasper Juel-Berg, and Berislav Tomicic, "Applying remote sensing to determine the percent imperviousness for urban drainage modelling," *Research Journal of Environmental Science*, vol. 12, no. 3, pp. 132–143, 2018.

[9] Weiqi Zhou, Jia Wang, and Mary L. Cadenasso, "Effects of the spatial configuration of trees on urban heat mitigation: A comparative study," *Remote Sensing of Environment*, vol. 195, pp. 1–12, 2017.

[10] Burr Settles, *Active Learning*, Synthesis Lectures on Artificial Intelligence and Machine Learning. Morgan and Claypool Publishers, 2012.

[11] Vivien Sainte Fare Garnot and Loïc Landrieu, "Metric-guided prototype learning," *CoRR*, vol. abs/2007.03047, 2020.

[12] Luca Bertinetto, Romain Mueller, Konstantinos Tertikas, Sina Samangooei, and Nicholas A Lord, "Making better mistakes: Leveraging class hierarchies with deep networks," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020, pp. 12506–12515.

[13] Romain Thoreau, Veronique Achard, Laurent Risser, Beatrice Berthelot, and Xavier Briottet, "Active learning for hyperspectral image classification: A comparative review," *IEEE Geoscience and Remote Sensing Magazine*, pp. 2–24, 2022.

[14] Wei Hu, Yangyu Huang, Li Wei, Fan Zhang, and Hengchao Li, "Deep convolutional neural networks for hyperspectral image classification," *Journal of Sensors*, vol. 2015, pp. 1–12, 07 2015.

[15] Nicolas Audebert, Bertrand Le Saux, and Sébastien Lefèvre, "Deep learning for classification of hyperspectral data: A comparative review," *IEEE geoscience and remote sensing magazine*, vol. 7, no. 2, pp. 159–173, 2019.