

COST-SENSITIVE MULTITASK ACTIVE LEARNING FOR CHARACTERIZATION OF URBAN ENVIRONMENTS WITH REMOTE SENSING

Christian Geiß¹, *Member, IEEE*, Matthias Thoma², and Hannes Taubenböck¹

¹ German Remote Sensing Data Center (DFD), German Aerospace Center (DLR), 82234 Wessling-Oberpfaffenhofen, Germany; christian.geiss@dlr.de, hannes.taubenboeck@dlr.de

² University of Applied Sciences Weihenstephan-Triesdorf, 85354 Freising, Germany; matthias.thoma@hswt.de

Abstract— In this letter, we propose a novel cost-sensitive multi-task active learning (CSMTAL) approach. Cost-sensitive active learning (CSAL) methods were recently introduced to specifically minimize labeling efforts emerging from ground surveys. Here, we build upon a CSAL method but compile a set of unlabeled samples from a learning set which can be considered relevant with respect to multiple target variables. To this purpose, a multi-task meta-protocol based on alternating selection is implemented. It comprises a so-called one-sided selection (i.e., single-task AL selection for a reference target variable with simultaneous labeling of the residual target variables) with a changing leading variable in an iterative selection process. Experimental results are obtained for the city of Cologne, Germany. The target variables to be predicted, using features from remote sensing and a Support Vector Machines framework, comprise “building type” and “roof type”. Comparative model accuracy evaluations underline the capability of the CSMTAL method to provide beneficial solutions with respect to a random sampling strategy and non-cost-sensitive multi-task active sampling.

Index Terms— Cost-sensitive Multi-task Active Learning, Support Vector Machines, Remote Sensing, Very High Resolution Imagery, LiDAR, Building Type, Roof Type

I. INTRODUCTION

The derivation of thematic information from remote sensing data is often based on techniques of empirical inference.

Such approaches foresee the collection of properly encoded prior knowledge (i.e., labeled samples) to infer a rule (e.g., a decision function) to accurately predict different thematic classes with respect to the residual instances under analyses (i.e., unlabeled samples). Thereby, the collection of a sufficient amount of prior knowledge can be very costly. Especially remote sensing-based applications determine frequently labor intensive ground truth surveys to compile a training set for model learning. Such applications comprise urban land use / land cover mapping [1],[2], retrieval of biophysical parameters [3], the quantitative description of urban morphology [4], or natural hazard-related vulnerability and risk assessments [5], [6], among numerous others. Different sampling strategies were followed to compile a proper training set (i.e., a pool of labeled samples) efficiently [7]–[9]. Thereby, active learning (AL) methods establish query functions to select unlabeled instances that can be considered to be the most valuable for improvement in terms of accuracy of a preliminary learned model. Then, selected unlabeled samples are labeled in a prioritized manner by a so-called oracle (e.g., an expert in the field). Thus, AL procedures aim

to compile effective training sets with few relevant labeled samples [10].

Recent works also consider labeling costs emerging from ground surveys [11], [12]. In this paper, we build upon the concept of cost-sensitive AL (CSAL), which internalizes the costs for labeling efforts with respect to sample selection. However, we render the AL problem as a multi-task (MTAL) optimization procedure. In a MTAL environment, the classification model is not strictly optimized for a single target variable. Instead, the goal is to learn an efficient solution with respect to multiple target variables simultaneously [13]. In the application context of this paper, we aim to characterize the urban environment. In particular, labels for multiple target variables must be assigned to a building in order to specify different properties (i.e., “building type”, and “roof type”) using features from remote sensing data. To this purpose, a MTAL meta-protocol is followed based on alternating selection (AS). The governing principle of AS is to implement a so-called one-sided selection (i.e., single-task AL selection for a reference target variable with simultaneous labeling of the residual target variables), with a changing leading variable in an iterative selection process. Given those methodological considerations, we introduce a novel method termed cost-sensitive multi-task active learning with alternating selection strategy (CSMTAL-AS). The concept of MTAL was recently employed by the authors of [14] in the context of remote sensing. There, MTAL was used to optimize so-called remote rapid visual screening assessments of buildings by an operator in a laboratory. As such, we build upon our previous work but establish here a cost-sensitive extension for optimization of data collection in ground surveys.

The remainder of the letter is organized as follows. Section II documents the developed CSMTAL-AS method. Section III is used to describe study area, data sets and parameterization of methods. Results of the actual experiments are reported in section IV. We give concluding remarks in section V.

II. COST-SENSITIVE MULTI-TASK ACTIVE LEARNING

The goal of the proposed CSMTAL-AS method is to select a batch $X = \{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_h\}$ of h samples from a pool of unlabeled samples U (i.e., learning set) to be included in the training set by evaluating three criteria: i) *uncertainty*, ii) *diversity*, and iii) *costs for labeling* with respect to a changing leading variable in an iterative selection process.

A. Uncertainty, Diversity, and Cost Criteria for AL

To incorporate uncertain samples, confidence $c(\mathbf{x})$ for each unlabeled sample $\mathbf{x} \in U$ is computed from its functional distance $f_i(\mathbf{x})$, $i = 1, 2, \dots, d$, to d decision functions of binary

Support Vector Machines (SVM) in a one-against-all or one-against-one multiclass architecture:

$$\begin{aligned} r_{1max} &= \arg \max_{i=1,2,\dots,d} \{f_i(\mathbf{x})\} \\ r_{2max} &= \arg \max_{j=1,2,\dots,d, j \neq r_{1max}} \{f_j(\mathbf{x})\} \\ c(\mathbf{x}) &= f_{r_{1max}}(\mathbf{x}) - f_{r_{2max}}(\mathbf{x}). \end{aligned} \quad (1)$$

The sample in the model which minimizes the numeric value of $c(\mathbf{x})$ is the sample closest to the decision boundary between the two most probable classes r_{1max} and r_{2max} and, thus, maximizes multiclass-level uncertainty (MCLU) [15].

To account for diversity in X , the angle-based diversity (ABD) method [16] quantifies similarity of uncertain samples. ABD is based on the cosine angle distance of hyperplanes which are induced by individual samples. Consequently, by relying on the kernel function $K(\cdot, \cdot)$, the cosine function of the (undirected) angles between two hyperplanes H_i and H_j , which are induced by unlabeled samples \mathbf{x}_i and \mathbf{x}_j (with normal vectors $\phi(\mathbf{x}_i)$ and $\phi(\mathbf{x}_j)$), is calculated as:

$$\begin{aligned} |\cos(\angle(H_i, H_j))| &= \frac{|\langle \phi(\mathbf{x}_i) \cdot \phi(\mathbf{x}_j) \rangle|}{\|\phi(\mathbf{x}_i)\| \|\phi(\mathbf{x}_j)\|} \\ &= \frac{K(\mathbf{x}_i, \mathbf{x}_j)}{\sqrt{K(\mathbf{x}_i, \mathbf{x}_i)K(\mathbf{x}_j, \mathbf{x}_j)}} \end{aligned} \quad (2)$$

and the angle between H_i and H_j :

$$\angle(H_i, H_j) = \cos^{-1} \left(\frac{K(\mathbf{x}_i, \mathbf{x}_j)}{\sqrt{K(\mathbf{x}_i, \mathbf{x}_i)K(\mathbf{x}_j, \mathbf{x}_j)}} \right). \quad (3)$$

Thus, the angle of two unlabeled samples indicates their diversity: it is small if they are similar and large if they are differing.

For introducing labeling costs in the model, different components are considered when compiling X . Those reflect the possible plurality of transportation modes (e.g., traveling by car or foot), which can be associated with different costs, and the actual time for assigning labels. Comparable to [11], total labeling costs, i.e., time for assigning a label to h samples of X at the l th iteration are quantified as:

$$t_l(X) = t_l^{mode1}(X) + t_l^{mode2}(X) + t_l^{labeling} \cdot h. \quad (4)$$

B. Cost-Sensitive Multi-Task Active Learning with Alternating Selection

We build upon a CSAL method as introduced by [11] with sequential forward selection. First, a set of uncertain candidates $m > h$ is drawn from U based on MCLU criterion. Subsequent to that, sets are compiled by choosing the most cheap and diverse samples from m uncertain samples. Here, a combined evaluation of the three criteria is established with a criterion function for the k th set $J_l(X_k)$:

$$J_l(X_k) = \lambda_{cost} t_l(X_k) + \lambda_{uncertainty} C_l(X_k) + \lambda_{diversity} D_l(X_k) \quad (5)$$

with $\lambda_{cost} + \lambda_{uncertainty} + \lambda_{diversity} = 1$; $C_l(X_k)$ is the average uncertainty of samples in the set and $D_l(X_k)$ represents the diversity. The latter is quantified by the average angle-based distances computed in kernel space between each pair of samples (equation 3) in X_k :

$$D_l(X) = \frac{2}{h \cdot (h-1)} \sum_{i=1}^{h-1} \sum_{j=i+1}^h \frac{K(\mathbf{x}_i, \mathbf{x}_j)}{\sqrt{K(\mathbf{x}_i, \mathbf{x}_i)K(\mathbf{x}_j, \mathbf{x}_j)}} \quad (6)$$

A small numeric value of $D_l(X)$ corresponds to a high diversity in the batch. The final batch is selected as that which minimizes the criterion function:

$$\min_{k=1,2,\dots,m} \{J_l(X_k)\}. \quad (7)$$

Analogous to previous work [11], a complete search is carried out by selecting each sample \mathbf{x}_i , $i = 1, 2, \dots, m$ from the uncertain m candidates. Each sample is sequentially added to the first position of the batch to determine the most favorable initial solution. Subsequent to that, further diverse and cheap samples are selected from m candidates and added to the residual positions of the batch by forward selection. Diversity and costs are weighted in a combined manner by parameter $\lambda_c = \frac{\lambda_{cost}}{\lambda_{cost} + \lambda_{diversity}}$. A high λ_c value favors cheap samples, whereas a low λ_c value causes a prioritized selection of diverse samples. A new sample \mathbf{x}_t is added to the batch at the l th iteration according to the optimization problem:

$$\mathbf{x}_t = \arg \min_{i=1,\dots,m} \left\{ \lambda_c t_l(X) + (1 - \lambda_c) \cdot \left[\max_{x_j \in X} \frac{K(\mathbf{x}_i, \mathbf{x}_j)}{\sqrt{K(\mathbf{x}_i, \mathbf{x}_i)K(\mathbf{x}_j, \mathbf{x}_j)}} \right] \right\}. \quad (8)$$

The previous steps are embedded in an MTAL meta-protocol based on *alternating selection*. The idea of AS is to implement one-sided selection with a changing leading variable in an iterative selection process [13]. A batch of uncertain, diverse and cheap unlabeled samples with respect to the leading variable is labeled intrinsically and the residual variables are labeled extrinsically within a single iteration of the model. The query function changes the leading variable by iterating through an ordered list of variables. Thereby, individual weights can be put upon the different variables by allowing certain variables to lead the selection more frequently than others (i.e., appear more often on the ordered list). However, here, we select the variables with equal weights. This corresponds to a round-robin strategy. In the experiments, we employ this CSMTAL *one-sided alternating selection – equally weighted* technique with a stacked feature vector from all variables for computing hyperplanes in kernel space regarding the ABD criterion [14]. An overview of the selection procedure for a single iteration is provided in Fig. 1.

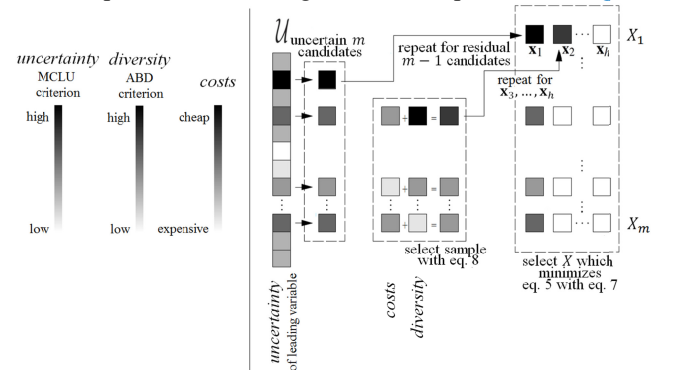


Fig. 1. Scheme of a single iteration of the CSMTAL-AS method for compilation of the batch.

The whole procedure to select unlabeled samples for inclusion in the training set is also described in the pseudocode of Algorithm 1.

Algorithm 1 CSMTAL-AS**Inputs:**

$\lambda_{\text{cost}}, \lambda_{\text{uncertainty}}, \lambda_{\text{diversity}}$ (weighting parameters that tune the tradeoff between cost, uncertainty and diversity)
 m (number of samples selected for each target variable based on their uncertainty)
 h (batch size)
 t (number of target variables)

Output:

X (set of unlabeled samples to be included in the training set)

1. Initialize $l = 1$.

Repeat

2. Initialize $k = 1$.
3. Compute $c(\mathbf{x})$ for each sample $\mathbf{x} \in U$ for currently leading target variable.
4. Select the set of m unlabeled samples with the lowest $c(\mathbf{x})$ value (most uncertain) $\{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_m\}$
5. Compute distance matrix with pairwise distances between all candidates, including current position.
- Repeat**
6. Initialize X_k to the empty set.
7. Include in X_k the k th most uncertain sample.
- Repeat**
8. Compute combination of diversity and cost with (8)
9. Include unlabeled sample \mathbf{x}_i which minimizes (8) in X .
- Until** $|X_k| = h$
10. $k = k + 1$
- Until** $k = n$
11. Select the batch X which minimizes (5) with (7).
12. Add labels to the set of samples $\{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_h\} \in X$ and include them in the current training set T .
13. $l = l + 1$
- if** $l > t$ **then** $l = 1$
- Until** stopping criterion

When the AL process is finished and a batch of previously unlabeled samples is labeled and moved from the learning set to the training set, the classification model is established with the considered SVM approach.

III. EXPERIMENTAL SETUP

A. Study Area and Data Sets

We simulate a ground survey with data from the city of Cologne, Germany. Building geometries were compiled from cadastral data sources, which were also made publicly available via the OpenStreetMap (OSM) project [17]. Building blocks for feature calculation were generated from the main street network of the OSM data set for Cologne. In addition, VHR multispectral imagery was deployed, which was acquired by the WorldView-II sensor. The pansharpened imagery features a spatial resolution of 50 cm. The deployed four-band multispectral (blue, green, red, NIR) imagery was acquired on January 31, 2014. In addition, elevation information from LiDAR data with a spatial resolution of 1 m was used [14]. Building characteristics for the complete building inventory of the city of Cologne are based on cadastral sources and were specifically compiled in the context of a flood-related decision support system. From this data set we selected the target variables “*building type*”, and “*roof type*” and rendered the experiment as a two-task annotation scenario. Target variables with affiliated classes and numbers of labeled samples are documented in Table I.

TABLE I

TARGET VARIABLES AND POPULATION OF THE CLASSES FOR TRAINING/LEARNING AND INDEPENDENT TEST SET

building type	train	test	Σ
<i>detached house</i>	12 549	3 463	16 012
<i>detached building block</i>	5 166	1 513	6 679
<i>perimeter block development</i>	10 275	2 871	13 146
<i>terraced house</i>	37 958	11 228	49 186
<i>garage</i>	39 646	4 132	43 778
Σ	105 594	23 207	128 801
roof type	train	test	Σ
<i>flat roof</i>	49 938	6 005	55 943
<i>single pitch roof</i>	9 100	1 325	10 425
<i>gable roof</i>	75 204	15 877	91 081
Σ	134 242	23 207	157 449

B. Experimental Setup

For a discriminative characterization of buildings, an exhaustive number of features was computed from three spatial levels (i.e., individual building geometries, aggregated building geometries and building blocks). Overall, 152 features were derived. Those features describe shape and extent of the buildings in two and three dimensions, are based on 1st and 2nd order spectral information, and characterize the spatial context and configuration, which the individual buildings are embedded in. A detailed list of the deployed features is provided in [14]. To compile compact feature sets, a multi-step feature selection procedure was carried out based on multivariate correlation analysis [14]. The final feature vectors comprise 20 dimensions for the variables *building type*, as well as *roof type*.

A batch size of $h = 5$ for AL strategies was adopted to allow for a feasible tradeoff between computational efforts and model accuracy. For the SVM, we deployed Gaussian RBF kernels $K(\mathbf{x}, \mathbf{x}_i) = \exp(-\gamma \|\mathbf{x} - \mathbf{x}_i\|^2)$. Tuning of C and γ was addressed by a grid search strategy based on fivefold cross-validation. Generalization accuracy was evaluated in terms of estimated κ statistic on the average of three independent trials with an exhaustive optimization of hyperparameters $\Phi \in \{C, \gamma\}$: $C = \{2^{-5}, 2^{-3}, \dots, 2^{15}\}$, $\gamma = \{2^{-15}, 2^{-13}, \dots, 2^3\}$. Hyperparameters were retuned very frequently at the beginning and less frequently at the end of the learning process [14]. Documented accuracies were computed from mean κ statistic values of 50 independent trials with initial training sets made up of two samples per class. Labeled samples were split into independent training and test sets, whereby a larger share of available samples was allocated to the training set. This was done to ensure that the spatial selection of samples is not biased with respect to the availability of potential training samples that can be queried in close spatial proximity. From the pool of labeled samples, a maximum of 800 samples were used for model learning and the complete pool of labeled test samples was used for accuracy evaluation.

Distances for the simulated ground survey were computed from the OSM street network. To this purpose, the street network was modeled as an undirected graph, where constituting points of street polygons were treated as knots and

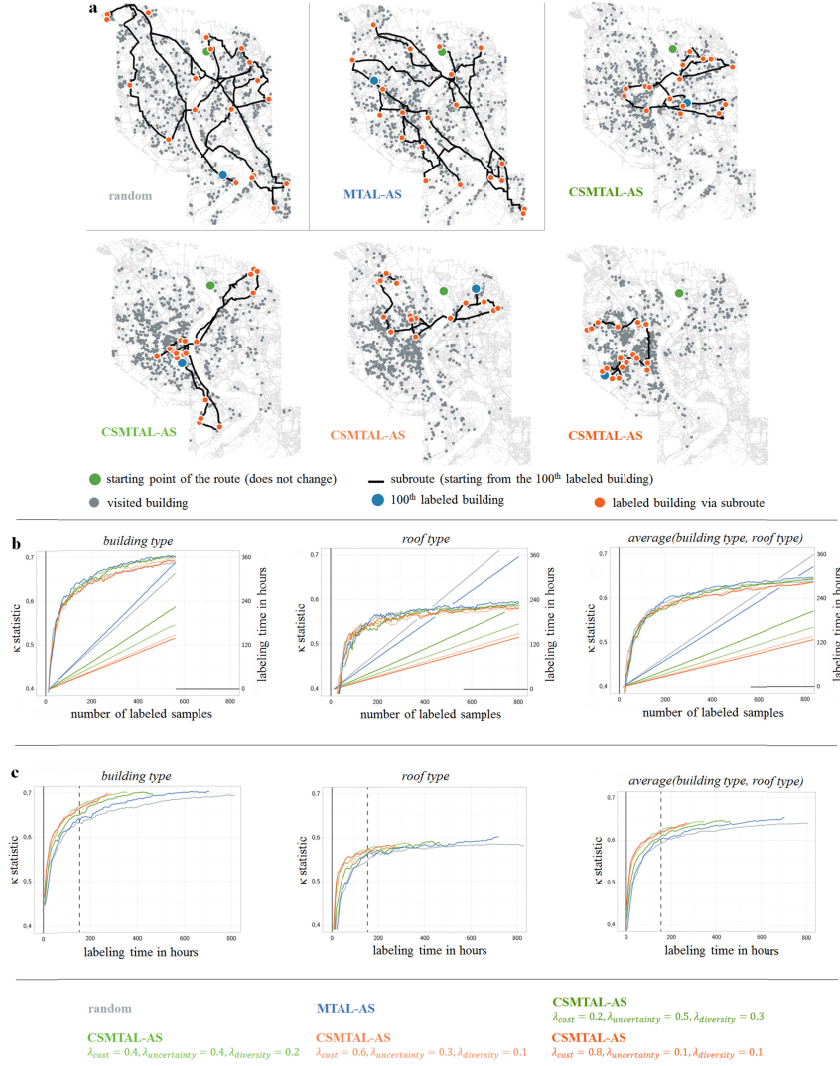


Fig. 2. Experimental results obtained for the city of Cologne. (a) Exemplary results from single model runs and affiliated routes for labeling of buildings. Subroutes, which start from the respective 100th labeled building, are highlighted for the subsequent 20 buildings for both the non-cost-sensitive benchmark methods (i.e., random selection and MTAL-AS), and the CSMTAL-AS method with different weights of the criterion function; (b) performance of the different methods as evaluated with κ statistic and labeling time for the two target variables and their average as a function of amount of prior knowledge (i.e., number of labeled samples); (c) performance of the different methods as evaluated with κ statistic for the two target variables and their average as a function of labeling time.

in-between distances were modeled as edge weights. We used the Dijkstra algorithm to compute the shortest distance between two knots in the graph [18]. To find the shortest path with respect to all buildings which are queried in a batch given the actual position of the surveyor within the graph, all pairwise distances of the buildings were computed first. Subsequent to that, the shortest path was determined with an algorithm with farthest-insertion-heuristic for solving problems of the traveling salesman [19]. The building which is visited last in a batch becomes the point of origin for the selection of the next batch in the subsequent AL model iteration. In the experiment, we modeled traveling efforts of the surveyor based on two modes: Based on the length of the shortest path to visit all buildings in a batch (i.e., $d_l^{car}(X)$), the traveling time by car was computed with $t_l^{car}(X) = d_l^{car}(X)/v^{car}$, whereby we assigned an inner-city average

traveling speed of $v^{car} = 25\text{km/h}$. Additional distances from the positions in the street network to a building were covered by foot with $v^{foot} = 5\text{km/h}$. Those distances must be covered twice (i.e., traveling from the car to the building and back): $t_l^{foot}(X) = 2 \cdot d_l^{foot}(X)/v^{foot}$. Actual *in situ* labeling efforts were quantified here with $t_l^{labeling} = 5\text{min}$ per building.

IV. EXPERIMENTAL RESULTS AND DISCUSSION

Results of experimental analysis are documented in Fig. 2. For comparative evaluation of the proposed approach, we also consider two non-cost-sensitive methods, i.e., a random labeling strategy and a MTAL method without cost constraint (i.e., MTAL-AS). To keep a concise letter, we did not follow potential alternative cost-sensitive multi-task methods such as cost-sensitive one-sided labeling, since it was shown that

purely one-sided selection strategies do not feature favorable performance properties in general [13], [14]. Four different results are provided for the CSMTAL-AS approach according to different weights of the criterion function, especially with respect to a consecutive increase of the cost criterion (i.e., $\lambda_{cost} = \{0.2, 0.4, 0.6, 0.8\}$).

Fig. 2a provides an overview on exemplary routes for labeling through the building inventory of Cologne, which were generated from the different methods. The starting point was fixed and subroutes, which start from the respective 100th labeled building, are visualized for the subsequent 20 buildings (i.e., labeling of four batches). If the cost criterion is weighted weakly (i.e., $\lambda_{cost} = \{0.2, 0.4\}$), buildings all over the urban area are queried. Thereby, subroutes feature an explorative nature, i.e., large distances are covered for labeling particularly interesting buildings. However, compared to random sampling and MTAL-AS, covered distances and affiliated timely resources are already cut in half. If the cost criterion is weighted more strongly (i.e., $\lambda_{cost} = \{0.6, 0.8\}$), time resources further decline by following more compact routes, which dominantly focus on inner-city areas. The large river Rhine divides the urban area of Cologne in eastern and western parts as indicated in Fig. 2a. Crossings from the eastern to the western parts of the city using the few bridges over the Rhine become less likely, and some parts of the urban area remain unvisited when weighting costs strongly.

Fig. 2b provides accuracy estimates for the two variables as well as their average as a function of amount of prior knowledge (i.e., number of labeled samples). The achieved accuracy levels underline the challenging nature of the classification problem. Thereby, best models encoded with a sufficient amount of prior knowledge achieve κ statistics up to 0.7 for the target variable “building type”, whereas models for the target variable “roof type” hardly feature substantial agreements (i.e., κ statistics exceeding values of 0.6). In addition to accuracy estimates, affiliated labeling times of the different approaches are provided. When estimating “building types”, MTAL-AS provides superior models in terms of accuracy, however, the CSMTAL strategies also allow for comparable accuracy levels. Thereby, an increased weighting of the costs consistently decreases accuracy, and accuracies of random sampling can hardly be exceeded with a highly weighted cost criterion. Nevertheless, when relating the achieved accuracy levels to the affiliated labeling times it can be seen that the CSMTAL approaches enable a beneficial tradeoff between labeling time and model accuracy. They allow for a considerable decrease of labeling time in concordance with only a slight decrease of accuracy. This tradeoff also becomes evident when inspecting results for the simultaneously labeled second target variable “roof type”. Again, MTAL-AS provides superior models in terms of accuracy but can only slightly exceed random selection. In parallel, the CSMTAL strategies feature slightly lower accuracy levels compared to MTAL-AS and random queries. However, the gap in terms of labeling efforts between the benchmark approaches and CSMTAL methods is large and clearly favors the latter. To enable further insights into the tradeoff between model accuracy and labeling time, the actual labeling time is plotted on the x-axis in Fig. 2c. There it can be seen that CSMTAL allows for substantially increased model

accuracy with the same labeling time compared to the non-cost-sensitive approaches.

V. CONCLUSIONS

In this letter, we have introduced a cost-sensitive multi-task AL approach for guided *in situ* data collection with respect to multiple target variables. To this purpose, we followed a strategy which foresees the collection of a batch of the most informative unlabeled samples based on a criterion function. The function internalizes an uncertainty, diversity, and cost criterion. This procedure is embedded in a multi-task meta-protocol, which is based on alternating selection of a leading variable in an iterative selection process. The proposed method was applied to the problem of optimization of data collection in a ground truth survey, in the context of classification of building characteristics using VHR multispectral imagery and LiDAR data. Experimental results underline the usefulness of the proposed query function, as it allows for substantially increased model accuracies given the same labeling time compared to non-cost-sensitive benchmark approaches.

REFERENCES

- [1] X. Huang, Q. Lu, and L. Zhang, “A multi-index learning approach for classification of high-resolution remotely sensed images over urban areas,” *ISPRS J. Photogramm. Remote Sens.*, vol. 90, no. 1, pp. 36–48, 2014.
- [2] C. Geiß, M. Klotz, A. Schmitt, and H. Taubenböck, “Object-based morphological profiles for classification of remote sensing imagery,” *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 10, pp. 5952–5963, Oct. 2016.
- [3] J. Verrelst, L. Alonso, G. Camps-Valls, J. Delegido, and J. Moreno, “Retrieval of vegetation biophysical parameters using Gaussian process techniques,” *IEEE Trans. Geosci. Remote Sens.*, vol. 50, no. 5, pt. 2, pp. 1832–1843, 2012.
- [4] H. Taubenböck *et al.*, “Delineation of central business districts in mega city regions using remotely sensed data,” *Remote Sens. Environ.*, vol. 136, pp. 386–401, Sep. 2013.
- [5] C. Geiß, P. Aravena Pelizari, M. Marconcini, W. Sengara, M. Edwards, T. Lakes, and H. Taubenböck, “Estimation of seismic buildings structural types using multi-sensor remote sensing and machine learning techniques,” *ISPRS J. Photogramm. Remote Sens.*, vol. 104, pp. 175–188, 2015.
- [6] C. Geiß, M. Milje, T. Lakes, and Taubenböck, H., “Estimation of Seismic Vulnerability Levels of Urban Structures With Multisensor Remote Sensing,” *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 9, no. 5, pp. 1913–1936, May 2016.
- [7] X. Huang, C. Weng, Q. Lu, T. Feng, and L. Zhang, “Automatic labelling and selection of training samples for high-resolution remote sensing image classification over urban areas,” *Remote Sens.*, vol. 7, no. 12, pp. 16024–16044, 2015.
- [8] S. Rougier, A. Puissant, A. Stumpf, and N. Lachiche, “Comparison of sampling strategies for object-based classification of urban vegetation from very high resolution satellite images,” *Int. J. Appl. Earth Observ. Geoinf.*, vol. 51, pp. 60–73, 2016.
- [9] D. Tuia, F. Ratle, F. Pacifici, M.F. Kanevski, W.J. Emery “Active Learning Methods for Remote Sensing Image Classification,” *IEEE Trans. Geosci. Remote Sens.*, vol. 47, no. 7, pp. Fig2218–2232, Jul. 2009.
- [10] 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 J. Sel. Topics Signal Process.*, vol. 5, no. 3, pp. 606–617, Jun. 2011.
- [11] B. Demir, L. Minello, and L. Bruzzone, “Definition of Effective Training Sets for Supervised Classification of Remote Sensing Images by a Novel Cost-Sensitive Active Learning Method,” *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 2, pp. 1272–1284, Feb. 2014.
- [12] C. Persello, B. Abdeslam, M. Dalponte, T. Gobakken, E. Næsset, and B. Schölkopf “Cost-Sensitive Active Learning with Lookahead: Optimizing Field Surveys for Remote Sensing Data Classification,” *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 10, pp. 6652–6664, Oct. 2014.
- [13] R. Reichart, K. Tomanek, U. Hahn, and A. Rappoport “Multi-task active learning for linguistic annotations,” *ACL-08: HLT*, pp. 861–869, 2008.
- [14] C. Geiß, M. Thoma, M. Pittore, M. Wieland, S. Dech, and H. Taubenböck, “Multitask Active Learning for Characterization of Built Environments with Multisensor Earth Observation Data,” *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 10, no. 12, pp. 5583–5597, Dec. 2017.
- [15] B. Demir, C. Persello, and L. Bruzzone, “Batch-Mode Active-Learning Methods for the Interactive Classification of Remote Sensing Images,” *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 3, pp. 1014–1031, Mar. 2011.
- [16] K. Brinker, “Incorporating diversity in active learning with support vector machines,” in *Proc. Int. Conf. Mach. Learn.*, 2003, pp. 59–66.
- [17] OpenStreetMap (Hrsg.), Geofabrik (Hrsg.): Download OpenStreetMap data for this region: Regierungsbezirk Köln. URL: <http://download.geofabrik.de/europe/germany/nordrhein-westfalen/koeln-regbez.html>. Last accessed: 28.09.2015
- [18] E.W. Dijkstra, “A Note on Two Problems in Connexion with Graphs,” *Numerische Mathematik*, vol. 1, no. 1, pp. 269–271, 1959.
- [19] D.J. Rosenkrantz, R.E. Stearns, P.M. Lewis, “An Analysis of Several Heuristics for the Traveling Salesman Problem,” *SIAM Journal on Computing*, vol. 6, no. 3, pp. 563–581, 1977.