Semantic Classification by Reasoning on the Whole Structure of Buildings using Statistical Relational Learning Techniques

Matteo Luperto, Alessandro Riva, and Francesco Amigoni

***Abstract*— Semantic mapping for autonomous mobile robots includes the place classification task that associates semantic labels (like 'corridor' or 'office') to rooms perceived in indoor environments. The mainstream approaches to place classification are characterized by *local* reasoning, where only features relative to the neighbourhood of each room are considered. In this paper, we propose a method for *global* reasoning on the whole structure of buildings, considered as single structured objects. We use a statistical relational learning algorithm, called *kLog*, and we compare it against a classifier, Extra-Trees, which resembles classical local approaches, in three tasks: classification of rooms, classification of entire floors of buildings, and validation of simulated worlds. Our results show that our global approach performs better than local approaches when the classification task involves reasoning on the regularities of buildings and when available information about rooms is coarse-grained.**

I. INTRODUCTION

Semantic mapping for autonomous mobile robots involves the task of associating semantic labels (like 'room' and 'corridor') to spatial entities perceived within an environment by means of sensors [1]. A semantic label identifies an entity, indicates its function, and allows a more informed interaction of the robots with the environment [2], [3]. One of the problems that are addressed in semantic mapping is that of *place classification* in buildings: identifying the presence, and labeling the function, of rooms or, more generally, spaces. Most of the methods for place classification follow an approach that, roughly, starts from the data perceived by sensors mounted on-board mobile robots (e.g., laser range scanners and cameras), extracts features from these data, and classifies the areas from which the data have been acquired using (supervised) machine learning techniques [4], [5].

These mainstream approaches for place classification are characterized by *local* reasoning, in which the input of the classifiers is a vector of features that refer to the space being classified or to its neighbourhood, following an attribute-value schema. These approaches have proven effective in classifying single rooms, but they do not naturally apply to more complex semantic classification tasks, like the classification of entire floors of buildings (i.e., assigning them labels like 'office' or 'school') since it is difficult to represent the structure of a building floor as a vector of features, being it better captured by a graph.

In this paper, we propose an approach to semantic classification that reasons on the whole structure of a building as a single entity. Each space of the environment is described not only by its geometrical features, but also by its role in the environment with respect to all the other spaces. To this end, we model indoor environments as graphs in which rooms are represented by nodes and edges represent doorways. Reasoning on such structured data is performed by exploiting *Statistical Relational Learning* (SRL) techniques. In recent years, the field of SRL has developed different techniques (like Markov Logic Networks [6]) for dealing and reasoning with complex multi-relational data in different settings [7]. The use of these techniques for semantic classification can enable a form of *global* reasoning on the structure of buildings, by considering their recurrent patterns. In particular, we originally apply to the context of semantic classification a state-of-the-art SRL algorithm, *kLog* [8], which is a method based on graph kernels for logical Entity/Relational structured data. Using kLog, we are able to effectively perform classification tasks similar to those performed by classical local methods, like labeling rooms, and also classification tasks that are not naturally addressed by local methods, because they require reasoning on structured knowledge, such as classifying an entire building according to its function and validating a simulated world. In order to evaluate (also quantitatively) our approach, we compare results obtained with it against those obtained with a standard attribute-value classification method, namely Extremely Randomized Trees (Extra-Trees) [9].

II. RELATED WORK

This section surveys a significant (although not exhaustive) sample of place classification systems presented in the literature.

In [10], a general framework for semantic mapping robots is presented. It is based on a multi-layered spatial representation of the environment and allows different input data and classifiers to be easily integrated. A similar system, with a hierarchical semantic map built from laser range scanner and camera data, is presented in [11].

In [12], the authors present a system which exploits multiple sensors (laser range scanners and cameras) in order to associate semantic labels to parts of the environment. Different features (like SIFT or CRFH for camera images) are extracted from sensory input collected in an environment. Three labels describing the environment are obtained independently by classifying the features with three different methods. A final label is calculated by merging the three labels with a multi-modal approach.

Matteo Luperto, Alessandro Riva, and Francesco Amigoni are with the Artificial Intelligence and Robotics Laboratory, Politecnico di Milano, Piazza Leonardo da Vinci 32, 20133 Milano, Italy {matteo.luperto, alessandro.riva, francesco.amigoni}@polimi.it

The system in [5] classifies single laser range scans and features extracted from cameras as belonging to rooms, corridors, hallways, or doorways using AdaBoost.

In [13], the authors propose a technique for place recognition and classification from image streams, called PLISS (Place Labeling through Image Sequence Segmentation). Differently from many supervised systems, PLISS can learn and update place models online and is able to operate even in the absence of training data.

The system presented in [4] uses a probabilistic framework that applies semantic labels starting from six kinds of environment features (objects, doorways, room shapes, room size, appearance, and associated spaces). The semantic map is represented as a probabilistic chain graph model that generalizes Bayesian Networks and Markov Random Fields. The use of a chain graph allows to account for the uncertainties of the sensory models, to classify a newly perceived room also according to the label of the rooms that are connected to it, and to predict the existence of a feature of a certain category (like a room and its label) in nearby unvisited space, extending the semantic map accordingly. A similar approach that uses a graph model based on associative Markov networks is that of [14]. Here, laser range scanner data are classified to label objects and rooms.

Another approach, reported in [15], integrates metric, topological, and semantic representations with information derived from natural language. It uses a factor graph formulation of the semantic properties and infers these properties by combining natural language descriptions and image- and laser-based scene classifications.

In most of the methods proposed in the literature, the association of semantic labels to perceived spaces is performed incrementally as sensor data are gathered and is independent of the semantic labels associated to previously seen spaces (with some relevant exceptions, like [4], that consider adjacent spaces). In this sense, we say that these methods are *local*, since each space is classified using only local features associated to the space itself, largely disregarding the structure of the whole map.

In [16] a different approach is introduced, which is focused on predicting the structure of a building. The authors consider a knowledge base of $38,000$ rooms (representing the MIT and KTH university campuses). Each floor of the buildings is represented as a graph, where nodes are rooms labeled according to their function (classrooms, offices, ...). Differently from the previous methods, that of [16] operates at an higher level of abstraction, reasoning directly on the rooms contained in the knowledge base, obtained from blueprints independently from the use of robots. In this sense, our approach is similar to that of [16], using graphs representing floorplans of buildings as source of knowledge. However, the focus of [16] is on predicting the structure of unseen environments rather than on their classification.

In this paper, we propose a method that, like [16], considers (and reasons on) the *whole* structure of a semantic map using techniques from the field of SRL [7], [17], that allow to reason with structured and multi-dimensional data. Markov Logic Networks (MLN) [6], Probabilistic Relational Models [18], and generalizations of Conditional Random Fields to arbitrary relations [19] have been used with success in several settings. Within SRL, logical representations are often used to provide a compact and expressive data format, such as in kLog [8], the SRL algorithm used in this paper and illustrated in Section III-C. Differently from local methods for semantic classification, which deal with uncertain sensor measurements, our approach only relies on a more abstract representation of buildings as graphs.

The use of SRL techniques in semantic mapping has been pioneered by [20], which proposes a method for obtaining semantic maps of indoor environments using a framework based on a MLN and data-driven Markov Chain Monte Carlo (MCMC) sampling. Using MCMC, the system samples many possible semantic worlds and selects the one that best fits the sensor data. The MLN evaluates the plausibility of each sampled semantic world, considering all the rooms (represented as atoms) at the same time, thus adopting a global approach. This approach shares some similarities with that presented in our paper, but in our case SRL techniques are used for classification purposes and not for evaluating a scoring function of a model.

III. A GLOBAL REASONING APPROACH

A. Problem Formulation

We represent (a floor of) an indoor environment as an undirected graph $G = (N, E)$, where each node $n \in N$ is a room and each edge $e = \{n, n'\} \in E$ (with $n, n' \in N$) represents a physical connection between two rooms n and n' (e.g., a doorway). A semantic label $\mathcal{L}(n)$ taken from a finite set of labels $\mathcal{L}$ is assigned to each node n. Semantic labels indicate the function of a room, such as 'corridor', 'office', or 'bathroom'. Each room n is also associated to a vector of features V_n which describe its geometrical structure. In our implementation, the vector V_n is composed of 8 geometrical features, computed from the polygon that approximates (as bounding box) the shape of the room n. Our features are a subset of those used in [21]: area, perimeter, area divided by perimeter, mean distance between the centroid and the shape boundary, form factor, circularity, normalized circularity, and average normalized distance between the centroid and the shape boundary.

As in [22]–[24], we exploit the concept of *building type*, developing specific models for each building type. Classes of buildings that have the same function, and thus similar structures, are called building types (e.g., schools, offices, ...). Reasoning on buildings of the same type allows to identify their common characteristics, that can be modeled effectively. We use two datasets of floorplans of buildings created by hand from labelled floorplans of real buildings (see [22], [24]) relative to two building types, SCHOOL and OFFICE. These two building types are selected because they are examples of common, large scale, and structured buildings, and are highly codified by regulations and design guidelines (compared, as example, to residential buildings which usually present more variety of possible structures).

The semantic classification tasks we address in this paper are defined as follows. The first task is place classification. Given a dataset $\mathcal{G}$ of labeled graphs representing buildings of a single building type, a function $f()$ that scores similarity between labeled graphs, and a query graph $\hat{G}$ whose nodes have associated vectors of features but not labels, the *place classification* task consists in finding the labels for the nodes in $\hat{G}$ such that $f(\hat{G}, \mathcal{G})$ is maximized.

Loosely speaking, local methods inductively build a model of the relation between the vectors of features V_n of nodes of graphs in $\mathcal{G}$ and the corresponding labels $\mathcal{L}(n)$ and, on the basis of this model, assigns a label $\mathcal{L}(\hat{n})$ to each node $\hat{n}$ of $\hat{G}$. Our global method, instead, inductively build a model of the labeled graphs in $\mathcal{G}$ and, on the basis of this model, assigns a label $\mathcal{L}(\hat{n})$ to each node $\hat{n}$ of $\hat{G}$.

The second task that we introduce in this paper is building classification. Assume to have a dataset $\mathcal{G}$ of graphs whose nodes are labeled as before and such that each graph $G \in \mathcal{G}$ has also a graph label L_G. Given such $\mathcal{G}$, a function $f()$ as before, and a query graph $\hat{G}$ whose nodes are labeled but whose graph label $L_{\hat{G}}$ is unknown, the *building classification* task consists in finding $L_{\hat{G}}$ such that $f(\hat{G}, \mathcal{G})$ is maximized. It is intuitive that building classification tasks involves recognizing recurrent patterns in graphs, which can be more naturally performed following a global approach.

B. Labeling Schemas

We use two hierarchically organized semantic labeling schemas to evaluate the impact of different sets of labels on semantic classification. The two schemas are common to all building types. At the top level, the labeling schema is called $\mathcal{L}_{R/C}$ and contains only two general categories (labels):

- ROOM: a space in which an activity is performed;
- CORRIDOR: a space that connects other spaces together.

In the second set of labels, called $\mathcal{L}_{F/C/E/S}$ = {FUNCTIONAL ROOM, CONNECTION, ENTRANCE, SERVICE ROOM}, CORRIDORs are specialized in:

- CONNECTION: a space that connects together different spaces within the same floor, such as corridors or hallways;
- ENTRANCE: a space that connects a floor to other floors or to the exterior of the building, such as an elevator or a staircase;

and ROOMs are specialized in:

- FUNCTIONAL ROOM: a space in which the core activity of the building is performed (e.g., a classroom in a school, an office or a meeting room in an office building);
- SERVICE ROOM: a space used to support the core activities of the building (e.g., restrooms or kitchens).

Note that $\mathcal{L}_{R/C}$ labels spaces according to coarse-grained information, while $\mathcal{L}_{F/C/E/S}$ uses a more fine-grained schema. Using current approaches in the literature, a $\mathcal{L}_{R/C}$ labeling can be performed from information stored in metric maps or from laser range scanner measurements, while labeling spaces according to a larger set of labels, such as $\mathcal{L}_{F/C/E/S}$, is more challenging and usually requires additional information (e.g., objects recognition to distinguish between functional and service rooms), as highlighted in [23]. This issue is further discussed in Section IV-B.

C. kLog

We use kLog [8] for performing semantic classification tasks by exploiting reasoning on the global structure of buildings. kLog input data are multiple relations between objects (possibly with attributes) represented in the form of an Entity/Relation knowledge base (E/R KB). This KB is expressed with a Prolog syntax and is composed of ground atoms under the closed world assumption. Entities and their relations are expressed using *signatures* and roughly correspond to ground atoms listed explicitly (extensional signatures) and to ground atoms implicitly defined using Prolog definite clauses (intensional signatures), respectively. Listing 1 shows a subset of the signatures we use, whose meaning is intuitive. Learning and classification tasks and queries are also expressed in Prolog syntax in a declarative way.

```
signature buildingtype(buildingtype::property)::extensional.
signature space(space_id::self)::extensional.
signature connected(s1::space, s2::space)::extensional.
signature label(space_id::space, label::property)::extensional.
signature area(space_id::space, area::property)::extensional.
signature perimeter(space_id::space, perimeter::property)::extensional.
signature iscorr(room_id::room)::intensional.
```

Listing 1: Examples of kLog signatures.

The label associated to a space is selected from one of the two sets of labels ($\mathcal{L}_{R/C}$ or $\mathcal{L}_{F/C/E/S}$), according to the task. Classification is performed by learning the *label(_, _)* relation or by learning the *buildingtype(_)* relation, according to the task. The former relation associates a label to a room, while the latter relation associates a label (the building type) to the query graph (see Listing 1).

```
interpretation(floor01,buildingtype(school)).
interpretation(floor01,space(s001)).
interpretation(floor01,label(s001,classroom)).
interpretation(floor01,space(s002)).
interpretation(floor01,label(s002,corridor)).
interpretation(floor01,space(s003)).
interpretation(floor01,label(s003,classroom)).
interpretation(floor01,connected(s001,s002)).
interpretation(floor01,connected(s002,s003)).
interpretation(floor01,area(s001,100)).
interpretation(floor01,perimeter(s001,40)).
...
```

Listing 2: Fragment of an interpretation.

kLog learns from interpretations. In our system, a set of instantiated signatures representing a floor of a building is called an *interpretation*. A fragment of an interpretation (called *floor01*) representing a floor of a school building composed of two classrooms connected by a corridor is reported in Listing 2.

Learning tasks, like binary or multi-class classifications, are performed in kLog using a Support Vector Machine (SVM) in a space defined using graph kernels. *Graph* *kernels* allow to operate in representative high-dimensional feature spaces without suffering the high cost of computing the feature spaces explicitly. Graph kernels used in kLog belong to the family of R-convolution kernels [25], based on decomposing structured data in smaller parts. Given a decomposition relation R^{-1} that extract from a graph G one of its possible decompositions in subgraphs $\{g\}$, the associated R-convolution kernel K compares all subgraphs of two graphs G and G', while a sub-kernel k_t compares the features $t \in \{1, 2, \ldots, T\}$ extracted from two subgraphs g and g':

$$K(G, G') = \sum_{g \in R^{-1}(G),\, g' \in R^{-1}(G')} \sum_{t=1}^{T} k_t(g, g')$$

Specifically, in kLog, the Neighborhood Subgraph Pairwise Distance Kernel (NSPDK) is used. NSPDK is extensively described in [8], [26]; here we just introduce the idea. The relation R^{-1} used by NSPDK extracts from a graph G all the pairs of neighborhood subgraphs g_i and g_j of radius r and centered on all possible pairs of nodes n_i and n_j that are at a distance d in the graph. The sub-kernel k_t in this case depends on r and d, let call it $k_{r,d}$, and computes the similarity of these pairs of subgraphs:

$$K(G, G') = \sum_r \sum_d \sum_{\substack{g_i, g_j \in R^{-1}(G) \\ g'_i, g'_j \in R^{-1}(G')}} k_{r,d}((g_i, g_j), (g'_i, g'_j))$$

For example, in an office building, the sub-kernel $k_{r,d}$ will assign large scores to symmetric portions of the building, since the r-neighborhoods of two symmetric nodes will be similar. In this sense, the use of graph kernels allows us to explicit capture the similarities between groups of nodes across a graph and between different graphs.

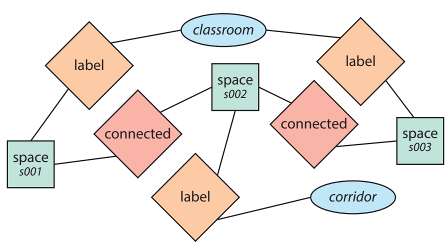


Fig. 1: Graphicalization of the Listing 2.

In kLog, the graph on which the graph kernel is applied is obtained from the E/R KB using a technique called *graphicalization*. Each entry found in the KB is converted into a node of a bipartite undirected graph whose nodes are ground atoms and edges connect an entity atom to a relationship atom if the identifier of the former appears as an argument in the latter. The graphicalization of the example of Listing 2 is shown in Fig. 1. Note that the generated graph embeds information about space connections (as the original graphs $\mathcal{G}$ of our dataset) plus other information (for example about labels: two spaces with the same label are now connected together via label nodes). Geometrical features V_n (i.e., area, perimeter, …of n) are represented in the graphicalization similarly to labels. NSPDK is then applied to the graphicalized KB and the similarity results it produces are processed by a standard SVM algorithm from the LibSVM repository[1] for performing classification. Please refer to [8] for further details.

D. Extremely Randomized Trees

In order to compare the reasoning on the whole structure of buildings with a local approach, we implement a state-of-the-art ensemble classification method, Extremely Randomized Tress (Extra-Trees) [9]. Extra-Trees algorithm builds an ensemble, or *forest*, of unpruned decision trees using a standard tree-based top-down classification procedures. Extra-Trees algorithm strongly randomizes on both attribute and cut-point choices while splitting a node of a tree. The two main differences with other tree-based ensemble methods are that (a) Extra-Trees algorithm splits nodes by choosing cut-points fully at random and (b) it uses the whole learning sample (rather than a bootstrap replica) to grow the trees. Extra-Trees are known to be computationally efficient, accurate, and robust for different classifications tasks.

Extra-Trees require an attribute-value representation of each room. Each room n is thus represented by its geometrical features V_n, to which its label $\mathcal{L}(n)$ is associated. (We also tried to add to V_n the labels of the rooms that are directly connected to n, but experimental results are similar.)

In order to test a mixed situation, in which global information is considered in a local setting, we use a second version of the Extra-Trees algorithm in which two more features that describe the role of a room n in the graph representing the building are added to V_n. These two features are related to *centrality*, which quantitatively evaluates the role of each node in the graph: high values of centrality correspond to most important nodes in the structure of the graph. In our context, centrality can be considered as a measure of the importance of each room within its floor structure. We use closeness and betweenness centrality. CLOSENESS centrality for a node n of a graph $G = (N, E)$ is defined according to the shortest distance between n and all other nodes of G:

$$\text{CLOSENESS}(n) = \sum_{t \in N \setminus \{n\}} 2^{-d_G(n,t)}$$

where $d_G(n, t)$ is the length of the shortest path on the graph G between the nodes n and t. BETWEENNESS centrality for a node n is defined as the number of shortest paths between two nodes $u \neq t (\neq n)$ that pass through n:

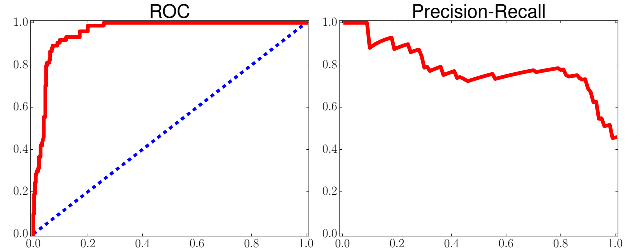
$$\text{BETWEENNESS}(n) = \sum_{u,t \in N, u \neq n \neq t} \frac{\tau_{ut}(n)}{\tau_{ut}}$$

where τ_{ut} is the total number of shortest paths from u to t and $\tau_{ut}(n)$ is the number of these paths that pass through n. This metric is normalized by $(|N| - 1)(|N| - 2)$. Classification tasks using Extra-Trees algorithm are evaluated both considering (Ex-T global) and not considering (Ex-T local) centrality among the features V_n of a room.

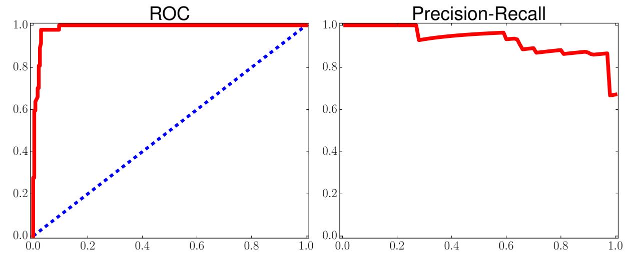
[1]https://www.csie.ntu.edu.tw/c̃jlin/libsvm/

	task	dataset	labels	error (%)
Ex-T local	●	S	RC	3.74
Ex-T global	●	S	RC	**2.44**
kLog	●	S	RC	6.82
Ex-T local	●	S	$FCES$	35.56
Ex-T global	●	S	$FCES$	**33.42**
kLog	●	S	$FCES$	46.89
Ex-T local	●	O	RC	7.21
Ex-T global	●	O	RC	**5.12**
kLog	●	O	RC	7.73
Ex-T local	●	O	$FCES$	36.15
Ex-T global	●	O	$FCES$	**32.80**
kLog	●	O	$FCES$	44.92
Ex-T local	■	S + O	RC	40.00
Ex-T global	■	S + O	RC	30.00
kLog	■	S + O	RC	**8.00**
Ex-T local	■	S + O	$FCES$	**5.00**
Ex-T global	■	S + O	$FCES$	**5.00**
kLog	■	S + O	$FCES$	8.00
Ex-T local	■	R + O	RC	16.67
Ex-T global	■	R + O	RC	8.33
kLog	■	R + O	RC	**0.00**
Ex-T local	■	R + O	$FCES$	8.33
Ex-T global	■	R + O	$FCES$	**0.00**
kLog	■	R + O	$FCES$	**0.00**

TABLE I: Results on different learning tasks using kLog and Extra-Trees (Ex-T local and global). Cyan circles stand for place classification tasks, blue squares for building classification tasks, and green squares for dataset validation tasks. S is the SCHOOL dataset, O is the OFFICE dataset, while R represents the RoboCup simulated office dataset. The labels column represents the label set used, while the error column indicates the classification error (best performance in bold).



(a) Classification of spaces using $\mathcal{L}_{R/C}$ for OFFICEs.



(b) Classification of spaces using $\mathcal{L}_{R/C}$ for SCHOOLs.

Fig. 2: Place classification for OFFICEs and SCHOOLs.

IV. EXPERIMENTS

In this section, we experimentally evaluate our global approach to semantic classification (based on kLog) and compare it against a classical local approach (Extra-Trees) on different tasks. All the results are obtained performing a leave-one-out cross validation over graphs of the two datasets relative to OFFICE and SCHOOL building types (each composed of 30 graphs of about 40 nodes each[2]). On a single core of a Intel Core i7-3610QM@2.30 GHz CPU with 8 GB RAM, all the kLog learning tasks discussed below run in less than 20 minutes (for all the datasets combined using cross validation), while Extra-Trees algorithm with a forest of 500 trees take approximatively slightly less then 60 minutes for performing the same tasks.

Evaluation is performed looking at the error rate of classification and at two curves. ROC (Receiver Operating Characteristic) curves plot the false positive rate on the x axis against the true positive rate on the y axis. Precision-recall curves plot the recall of classification on x axis against the precision of classification on y axis. Both axes go from 0 to 1 and the quality of classification is given by the area underlying the curve. Area equal to 1 means a “perfect classifier” while the line with angle $\pi/4$ (area equal to 0.5) represents the performance of a random classifier.

[2]Datasets are available upon request.

A. Place Classification

The first learning task mimics the classical task of place classification of rooms n of an environment, knowing the features V_n obtained from sensor readings, similarly to [21] (see Section III-A). In this setting, we first consider the top level $\mathcal{L}_{R/C}$ labeling schema, asking kLog to classify each space as ROOM or CORRIDOR (discriminating between the two labels is of interest in tasks like exploration [2], [3]). In other words, it performs a binary classification task for each space in the interpretation representing $\hat{G}$.

Fig. 2 shows the ROC and the precision-recall curves obtained by kLog for the building types OFFICE and SCHOOL. kLog is able to classify correctly ROOMs and CORRIDORs, as the large areas under the curves suggest. However, in both cases, Extra-Trees perform slightly better than kLog, as shown in Table I. We can explain these results considering that it is easy to distinguish between ROOMs and CORRIDORs within a specific building type using only local features (e.g., the number of doors and the shape, usually squared for rooms and rectangular for corridors [22], [23]). This trend is confirmed when we use the larger set of labels $\mathcal{L}_{F/C/E/S}$, where the robustness (due to its randomness) of Extra-Trees outperforms the less flexible SVM of kLog. It is interesting to point out that the use of global knowledge in form of centrality features improves the results of Extra-Trees. The error is rather high when using $\mathcal{L}_{F/C/E/S}$, probably because geometrical features alone are not sufficient to discriminate between these types of rooms and additional information is required (as in [27]).

Overall, our results show that the proposed global approach based on kLog is able to perform place classification with performance comparably similar to that of a classical attribute-value local approach. Outcomes of a place classification task performed by kLog could be integrated in a

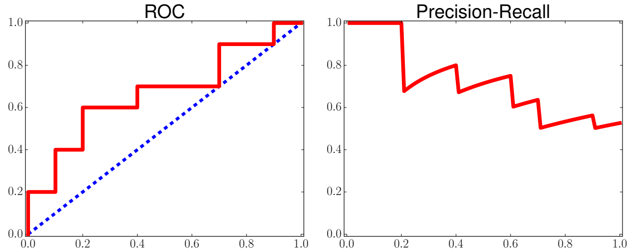


Fig. 3: Building classification in kLog using $\mathcal{L}_{R/C}$.

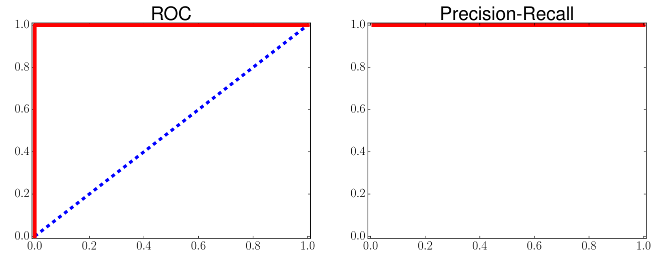


Fig. 4: Validation of structural realism of simulated worlds using kLog.

standard semantic mapping approach. For example, it is in principle easy to extend a framework like that of [4], which assigns labels according to different features derived from camera data, to include also a new feature based on global reasoning, which could propose a new label not based on sensorial data, but on the global structure of the portion of a building which is currently known.

B. *Building Classification*

Classifying entire buildings according to their function is a task that is much less explored in the literature, but that is intuitively expected to require reasoning about regularities and patterns of a specific building type. We refer to this task as *building classification*: we assume to know the structure, the geometry, and the room labels of a (floor of a) building, and we ask kLog and Extra-Trees to perform binary classification categorizing the function of the building as OFFICE or SCHOOL (see Section III-A). As input data we use the datasets (all the interpretations of) OFFICE and SCHOOL combined together. Since Extra-Trees reason only on local data and can classify only a room at a time, we ask the classifier to label each room according to a building type, and then we take the most frequent label to classify the whole floor. We perform tests using $\mathcal{L}_{R/C}$ and $\mathcal{L}_{F/C/E/S}$ schemas for labeling the rooms of the initial interpretations. Using $\mathcal{L}_{R/C}$ we assume that only a limited knowledge on the function of the rooms that a robot has viewed is available, while using $\mathcal{L}_{F/C/E/S}$ we assume to have more detailed knowledge on the function of the spaces.

Fig. 3 shows the ROC and the precision-recall curves for kLog when using $\mathcal{L}_{R/C}$. Good results are obtained, showing that, when the task is the classification of the whole building, reasoning on its global structure can compensate for a limited amount of information about the functions of the single spaces. If we compare the kLog results to those obtained using Extra-Trees (Table I) we can see that kLog outperforms Extra-Trees. Since Extra-Trees are usually considered more robust than SVMs, the better performance of kLog can be traced back to its exploitation of global relationships between parts of the building. However, using $\mathcal{L}_{F/C/E/S}$, a more informative knowledge on the function of the spaces (which, as discussed in Section III-B, could be in principle more difficult to obtain and more error-prone) overshadows the global properties of the environment, allowing Extra-Trees to carry out more accurate classifications. kLog performs better with the $\mathcal{L}_{R/C}$ labeling of rooms, namely when coarse-grained information about rooms is available. This can happen, for instance, when a robot can only use data from laser range scanners (which can be enough to distinguish between rooms and corridors, as shown [28] and discussed in [23]) and data from cameras (which could be used to distinguish between a more detailed set of labels like $\mathcal{L}_{F/C/E/S}$ as done, for example, in [4]) are not available, e.g., due to camera failure.

The ability of classifying whole buildings could be useful for a robot moving in large heterogeneous structured environments, like a university campus, that needs to know whether the floor in which it is currently navigating hosts teaching activities (i.e., it is composed of classrooms and lecture halls) or the administrative section of the campus (i.e., it is composed of faculty's offices, reception, ...) in order to make more informed decisions. We further test this ability by using the publicly available datasets of the MIT campus [16]. We randomly select a sample of 33 floorplans belonging to 3 different types of buildings: DORMs, OFFICEs, and RESEARCH LABs, and we use kLog to perform a (leave-one-out cross validation) classification with the original labeling schema of the datasets, obtaining a $93,94\%$ accuracy on assigning the correct type to the floorplans, which corresponds to a very accurate classification.

C. *Validation of Simulated Worlds*

The use of simulations in robotics has become increasingly relevant in the last years, especially as a way to preliminarily test systems that will be later assessed in real-world settings. If simulation tools have become more and more *physically realistic* [29], the design of simulated environments for mobile robots is usually receiving less attention, for example with the consequence that not all simulated buildings are realistic (i.e., are similar to their real counterparts) from a structural point of view. This situation can hinder the transfer of results from simulation to the real world. In this section we propose a method, based on our SRL approach, to test if a simulated environment presents the same structural characteristics of a similar real-world environment. In the context of our approach, this can be done by evaluating if the topological and geometrical properties of spaces of a (floor of a) simulated building are coherent to those of (floors of a) real-world buildings of the same type. In this way, we can assess if a simulated environment is *structurally realistic*.

We build a dataset of all the office environments used during the finals of the 2012 and 2013 editions of the Virtual Robot Competition of the Robocup Rescue Simulation League [30], which includes 6 simulated environments, designed to test the behavior of robots in an Urban Search and Rescue task. In kLog and Extra-Trees, validation of such dataset can be represented as a binary classification of interpretations (similar to a building classification), in which each interpretation is categorized as a positive instance of OFFICE or as negative example of an unrealistic simulation, using as input data all the interpretations of OFFICE and of RoboCup simulated offices combined together. Since Extra-Trees do not allow structured prediction, we consider the most frequent label, as for the building classification task. Fig. 4 shows that kLog is always able to distinguish a real-world office building from an unrealistic office building simulated in the competition, outperforming Extra-Trees, as shown in Table I. It is interesting to point out that, in this specific task, centrality degrades Extra-Trees performance and a truly global approach seems required.

V. CONCLUSIONS

This paper has discussed how SRL techniques can be used for reasoning on the whole structure of buildings exploiting information on the connections and on the geometry of rooms, in the context of semantic classification of spaces. Three applications assessed the potential of our approach, namely classification of rooms, classification of buildings, and validation of simulated worlds. The outcomes show that using structured learning methods for semantic classification is promising. A global approach is especially useful when the task requires capturing and reasoning on the regularities of buildings and the available information about room labels is coarse-grained. While semantic classification usually requires knowledge about the environment geometry and appearance (obtained from sensors), the use of knowledge about the whole structure of the building could reduce the amount of such information that is required. So, our results on the use of kLog suggest a trade-off between knowledge of details of single rooms and knowledge of the whole structure of a building when performing semantic classification.

Future work includes the further assessment of the proposed approach, with applications to other classification tasks and the extension of the semantic representation of buildings to involve multiple relations between nodes (e.g., explicitly identifying and representing symmetries between parts of a building). Finally, our structured learning method can be integrated in a standard semantic mapping framework to provide a new perspective on the building structure that can be used to produce better semantic maps.

REFERENCES

[1] I. Kostavelis and A. Gasteratos, "Semantic mapping for mobile robotics tasks: A survey," *Robot Auton Syst*, vol. 66, pp. 86–103, 2015.

[2] A. Quattrini Li, R. Cipolleschi, M. Giusto, and F. Amigoni, "A semantically-informed multirobot system for exploration of relevant areas in search and rescue settings," *Auton Robot*, vol. 40, no. 4, pp. 581–597, 2016.

[3] C. Stachniss, O. Mozos, and W. Burgard, "Speeding-up multi-robot exploration by considering semantic place information," in *Proc. ICRA*, 2006, pp. 1692–1697.

[4] A. Pronobis and P. Jensfelt, "Large-scale semantic mapping and reasoning with heterogeneous modalities," in *Proc. ICRA*, 2012, pp. 3515–3522.

[5] O. Mozos, R. Triebel, P. Jensfelt, A. Rottmann, and W. Burgard, "Supervised semantic labeling of places using information extracted from sensor data," *Robot Auton Syst*, vol. 55, no. 5, pp. 391–402, 2007.

[6] M. Richardson and P. Domingos, "Markov logic networks," *Mach Learn*, vol. 62, no. 1-2, pp. 107–136, 2006.

[7] L. Getoor and B. Taskar, *Introduction to Statistical Relational Learning*. MIT Press, 2007.

[8] P. Frasconi, F. Costa, L. De Raedt, and K. De Grave, "klog: A language for logical and relational learning with kernels," *Artif Intell*, vol. 217, pp. 117–143, 2014.

[9] P. Geurts, D. Ernst, and L. Wehenkel, "Extremely randomized trees," *Mach Learn*, vol. 63, no. 1, pp. 3–42, 2006.

[10] H. Zender, O. Mozos, P. Jensfelt, G. Kruijff, and W. Burgard, "Conceptual spatial representations for indoor mobile robots," *Robot Auton Syst*, vol. 56, no. 6, pp. 493–502, 2008.

[11] C. Galindo, A. Saffiotti, S. Coradeschi, P. Buschka, J. Fernandez-Madrigal, and J. González, "Multi-hierarchical semantic maps for mobile robotics," in *Proc. IROS*, 2005, pp. 2278–2283.

[12] A. Pronobis, O. Mozos, B. Caputo, and P. Jensfelt, "Multi-modal semantic place classification," *Int J Robot Res*, vol. 29, no. 2-3, pp. 298–320, 2010.

[13] A. Ranganathan, "PLISS: Detecting and labeling places using online change-point detection," in *Proc. RSS*, 2010.

[14] R. Triebel, R. Schmidt, O. Mozos, and W. Burgard, "Instance-based AMN classification for improved object recognition in 2D and 3D laser range data," in *Proc. IJCAI*, 2007, pp. 2225–2230.

[15] S. Hemachandra, M. Walter, S. Tellex, and S. Teller, "Learning spatial-semantic representations from natural language descriptions and scene classifications," in *Proc. ICRA*, 2014, pp. 2623–2630.

[16] A. Aydemir, P. Jensfelt, and J. Folkesson, "What can we learn from 38,000 rooms? Reasoning about unexplored space in indoor environments," in *Proc. IROS*, 2012, pp. 4675–4682.

[17] L. De Raedt, *Logical and Relational Learning*. Springer, 2008.

[18] N. Friedman, L. Getoor, D. Koller, and A. Pfeffer, "Learning probabilistic relational models," in *Proc. IJCAI*, 1999, pp. 1300–1309.

[19] C. Sutton and A. McCallum, "An introduction to conditional random fields for relational learning," in *Introduction to Statistical Relational Learning*, L. Getoor, Ed. MIT Press, 2006, pp. 93–128.

[20] Z. Liu and G. von Wichert, "A generalizable knowledge framework for semantic indoor mapping based on Markov logic networks and data driven MCMC," *Future Gener Comp Sy*, vol. 36, pp. 42–56, 2014.

[21] O. Martinez Mozos, *Semantic Labeling of Places with Mobile Robots*. Springer, 2010.

[22] M. Luperto, A. Quattrini Li, and F. Amigoni, "A system for building semantic maps of indoor environments exploiting the concept of building typology," in *Proc. RoboCup*, 2013, pp. 504–515.

[23] M. Luperto and F. Amigoni, "Exploiting structural properties of buildings towards general semantic mapping systems," in *Proc. IAS-13*, 2014, pp. 375–387.

[24] M. Luperto, L. D'Emilio, and F. Amigoni, "A generative spectral model for semantic mapping of buildings," in *Proc. IROS*, 2015, pp. 4451–4458.

[25] T. Gärtner, J. W. Lloyd, and P. Flach, "Kernels and distances for structured data," *Mach Learn*, vol. 57, no. 3, pp. 205–232, 2004.

[26] F. Costa and K. De Grave, "Fast neighborhood subgraph pairwise distance kernel," in *Proc. ICML*, 2010, pp. 255–262.

[27] A. Rottmann, O. Mozos, C. Stachniss, and W. Burgard, "Semantic place classification of indoor environments with mobile robots using boosting," in *Proc. AAAI*, 2005, pp. 1306–1311.

[28] O. Mozos, C. Stachniss, and W. Burgard, "Supervised learning of places from range data using AdaBoost," in *Proc. ICRA*, 2005, pp. 1730–1735.

[29] S. Carpin, M. Lewis, J. Wang, S. Balakirsky, and C. Scrapper, "USARSim: A robot simulator for research and education," in *Proc. ICRA*, 2007, pp. 1400–1405.

[30] "RoboCup Rescue League," http://wiki.robocup.org/wiki/Rescue_Simulation_League, 2016, [Online; accessed 7-September-2016].