
Collective Classification for Labeling of Places and Objects in 2D and 3D Range Data

Rudolph Triebel¹, Óscar Martínez Mozos², and Wolfram Burgard²

¹ Autonomous Systems Lab, ETH Zürich, Switzerland
rudolph.triebel@mavt.ethz.ch

² Department of Computer Science, University of Freiburg, Germany
{omartine,burgard}@informatik.uni-freiburg.de

Abstract. In this paper, we present an algorithm to identify types of places and objects from 2D and 3D laser range data obtained in indoor environments. Our approach is a combination of a collective classification method based on associative Markov networks together with an instance-based feature extraction using nearest neighbor. Additionally, we show how to select the best features needed to represent the objects and places, reducing the time needed for the learning and inference steps while maintaining high classification rates. Experimental results in real data demonstrate the effectiveness of our approach in indoor environments.

1 Introduction

One key application in mobile robotics is the creation of geometric maps using data gathered with range sensors in indoor environments. These maps are usually used for navigation and represent free and occupied spaces. However, whenever the robots are designed to interact with humans, it seems necessary to extend these representations of the environment to improve the human-robot communication. In this work, we present an approach to extend indoor laser-based maps with semantic terms like “corridor”, “room”, “chair”, “table”, etc, used to annotate different places and objects in 2D or 3D maps. We introduce the instance-based associative Markov network (iAMN), which is an extension of associative Markov networks together with instance-based nearest neighbor methods. The approach follows the concept of collective classification in the sense that the labeling of a data point in the space is partly influenced by the labeling of its neighboring points. iAMNs classify the points in a map using a set of features representing these points. In this work, we show how to choose these features in the different cases of 2D and 3D laser scans. Experimental results obtained in simulation and with real robots demonstrate the effectiveness of our approach in various indoor environments.

2 Related Work

Several authors have considered the problem of adding semantic information to 2D maps. Koenig and Simmons (1998) apply a pre-programmed routine to detect doorways. Althaus and Christensen (2003) use sonar data to detect corridors and doorways. Moreover, Friedman et al. (2007) introduce Voronoi random fields as a technique for mapping the topological structure of indoor environments. Finally, Martinez Mozos et al. (2005) use AdaBoost to create a semantic classifier to classify free cells in occupancy maps.

Also the problem of recognizing objects from 3D data has been studied intensively. Osada et al. (2001) propose a 3D object recognition technique based on shape distributions. Additionally, Huber et al. (2004) present an approach for parts-based object recognition. Boykov and Huttenlocher (1999) propose an object recognition method based on Markov random fields. Finally, Anguelov et al. (2005) present an associative Markov network approach to classify 3D range data. This paper is based on our previous work (Triebel et al. (2007)) which introduces the instance-based associative Markov networks.

3 Collective Classification

In most standard spatial classification methods, the label of a data point only depends on its local features but not on the labeling of nearby data points. However, in practice one often observes a statistical dependence of the labeling associated to neighboring data points. Methods that use the information of the neighborhood are denoted as *collective classification* techniques. In this work, we use a collective classifier based on associative Markov networks (AMNs) (Taskar et al. (2004)), which is improved with an instance-based nearest-neighbor (NN) approach.

3.1 Associative Markov Networks

An associative Markov network is an undirected graph in which the nodes are represented by N random variables y_1, \dots, y_N . In our case, these random variables are discrete and correspond to the semantic label of each of the data points $\mathbf{p}_1, \dots, \mathbf{p}_N$, each represented by a vector $\mathbf{x}_i \in \mathbb{R}^L$ of local features. Additionally, edges have associated a vector \mathbf{x}_{ij} of features representing the relationship between the corresponding nodes. Each node y_i has an associated non-negative potential $\varphi(\mathbf{x}_i, y_i)$. Similarly, each edge (y_i, y_j) has a non-negative potential $\psi(\mathbf{x}_{ij}, y_i, y_j)$ assigned to it. The node potentials reflect the fact that for a given feature vector \mathbf{x}_i some labels are more likely to be assigned to \mathbf{p}_i than others, whereas the edge potentials encode the interactions of the labels of neighboring nodes given the edge features \mathbf{x}_{ij} . Whenever the potential of a node or edge is high for a given label y_i or a label pair (y_i, y_j) ,

the conditional probability of these labels given the features is also high. The conditional probability that is represented by the network is expressed as:

$$P(\mathbf{y} | \mathbf{x}) = \frac{1}{Z} \prod_{i=1}^N \varphi(\mathbf{x}_i, y_i) \prod_{(ij) \in E} \psi(\mathbf{x}_{ij}, y_i, y_j), \quad (1)$$

where the partition function $Z = \sum_{\mathbf{y}'} \prod_{i=1}^N \varphi(\mathbf{x}_i, y'_i) \prod_{(ij) \in E} \psi(\mathbf{x}_{ij}, y'_i, y'_j)$.

The potentials can be defined using the log-linear model proposed by Taskar et al. (2004). However, we use a modification of this model in which a weight vector $\mathbf{w}^k \in \mathbb{R}^{d_n}$ is introduced for each class label $k = 1, \dots, K$. Additionally, a different weight vector $\mathbf{w}_e^{k,l}$, with $k = y_i$ and $l = y_j$ is assigned to each edge. The potentials are then defined as:

$$\log \varphi(\mathbf{x}_i, y_i) = \sum_{k=1}^K (\mathbf{w}_n^k \cdot \mathbf{x}_i) y_i^k \quad (2)$$

$$\log \psi(\mathbf{x}_{ij}, y_i, y_j) = \sum_{k=1}^K \sum_{l=1}^K (\mathbf{w}_e^{k,l} \cdot \mathbf{x}_{ij}) y_i^k y_j^l, \quad (3)$$

where y_i^k is an indicator variable which is 1 if point \mathbf{p}_i has label k and 0, otherwise. In a further refinement step in our model, we introduce the constraints $\mathbf{w}_e^{k,l} = \mathbf{0}$ for $k \neq l$ and $\mathbf{w}_e^{k,k} \geq \mathbf{0}$. This results in $\psi(\mathbf{x}_{ij}, k, l) = 1$ for $k \neq l$ and $\psi(\mathbf{x}_{ij}, k, k) = \lambda_{ij}^k$, where $\lambda_{ij}^k \geq 1$. The idea here is that edges between nodes with different labels are penalized over edges between equally labeled nodes.

If we reformulate Equation 1 as the conditional probability $P_{\mathbf{w}}(\mathbf{y} | \mathbf{x})$, where the parameters ω are expressed by the weight vectors $\mathbf{w} = (\mathbf{w}_n, \mathbf{w}_e)$, and plugging in Equations (2) and (3), we then obtain that $\log P_{\mathbf{w}}(\mathbf{y} | \mathbf{x})$ equals

$$\sum_{i=1}^N \sum_{k=1}^K (\mathbf{w}_n^k \cdot \mathbf{x}_i) y_i^k + \sum_{(ij) \in E} \sum_{k=1}^K (\mathbf{w}_e^{k,k} \cdot \mathbf{x}_{ij}) y_i^k y_j^k - \log Z_{\mathbf{w}}(\mathbf{x}). \quad (4)$$

In the learning step we try to maximize $P_{\mathbf{w}}(\mathbf{y} | \mathbf{x})$ by maximizing the margin between the optimal labeling $\hat{\mathbf{y}}$ and any other labeling \mathbf{y} (Taskar et al. (2004)). This margin is defined by:

$$\log P_{\omega}(\hat{\mathbf{y}} | \mathbf{x}) - \log P_{\omega}(\mathbf{y} | \mathbf{x}). \quad (5)$$

The inference in the unlabeled data points is done by finding the labels \mathbf{y} that maximize $\log P_{\mathbf{w}}(\mathbf{y} | \mathbf{x})$. We refer to Triebel et al. (2007) for more details.

3.2 Instance-Based AMNs

The main drawback of the AMN classifier explained previously, which is based on the log-linear model, is that it separates the classes linearly. This assumes

that the features are separable by hyper-planes, which is not justified in all applications. This does not hold for instance-based classifiers such as the nearest-neighbor (NN), in which a query data point $\tilde{\mathbf{p}}$ is assigned to the label that corresponds to the training data point \mathbf{p} whose features \mathbf{x} are closest to the features $\tilde{\mathbf{x}}$ of $\tilde{\mathbf{p}}$. In the learning step, the NN classifier simply stores the entire training data set and does not compute a reduced set of training parameters.

To combine the advantage of instance-based NN classification with the AMN approach, we convert the feature vector $\tilde{\mathbf{x}}$ of the query point $\tilde{\mathbf{p}}$ using the transform $\tau: \mathbb{R}^L \rightarrow \mathbb{R}^K: \tau(\tilde{\mathbf{x}}) = (d(\tilde{\mathbf{x}}, \hat{\mathbf{x}}_1), \dots, d(\tilde{\mathbf{x}}, \hat{\mathbf{x}}_K))$, where K is the number of classes and $\hat{\mathbf{x}}_k$ denotes the training example with label k closest to $\tilde{\mathbf{x}}$. The transformed features are more easily separable by hyperplanes. Additionally, the N nearest neighbors can be used in the transform function.

4 Feature Extraction in 2D maps

In this paper, indoor environments are represented by two dimensional occupancy grid maps (Moravec (1988)). The unoccupied cells of a grid map form an 8-connected graph which is used as the input to the iAMN. Each cell is represented by a set of single-valued geometrical features calculated from the 360° laser scan in that particular cell as shown by Martínez Mozos et al. (2005).

Three dimensional scenes are presented by point clouds which are extracted with a laser scan. For each 3D point we computed *spin images* (Johnson (1997)) with a size of 5×10 bins. The spherical neighborhood for computing the spin images had a radius between 10 and 15cm, depending on the resolution of the input data.

5 Feature Selection

One of the problems when classifying points represented by range data consists in selecting the size L of the features vectors \mathbf{x} . The number of possible features that can be used to represent each data point is usually very large and can easily be in the order of hundreds. This problem is known as *curse of dimensionality*. There are at least two reasons to try to reduce the size of the feature vector. The most obvious one is the computational complexity, which in our case, is also the more critical. We have to learn an inference in networks with thousands of nodes. Another reason is that although some features may carry a good classification when treated separately, maybe there is a little gain if they are combined together if they have a high mutual correlation (Theodoridis and Koutroumbas (2006)).

In our approach, the size of the feature vector for 2D data points is of the order of hundreds. The idea is to reduce the size of the feature vectors when used with the iAMN and at the same time try to maintain their class

discriminatory information. To do this we apply a scalar feature selection procedure which uses a class separability criterion and incorporates correlation information. As separability criterion C , we use the Fisher's discrimination ratio (FDR) extended to the multi-class case (Theodoridis and Koutroumbas (2006)). For a scalar feature f and K classes $\{w_1, \dots, w_K\}$, $C(f)$ can be defined as:

$$C(f) = FDR_f = \sum_i^K \sum_{j \neq i}^K \frac{(\mu_i - \mu_j)^2}{\sigma_i + \sigma_j}, \quad (6)$$

where the subscripts i, j refer to the mean and variance of the classes w_i and w_j respectively. Additionally, the cross-correlation coefficient between any two features f and g given T training examples is defined as:

$$\rho_{fg} = \frac{\sum_{t=1}^T x_{tf} x_{tg}}{\sqrt{\sum_{t=1}^T x_{tf}^2 \sum_{t=1}^T x_{tg}^2}}, \quad (7)$$

where x_{tf} denotes the value of the feature f in the training example t . Finally, the selection of the best L features involves the following steps:

- Select the first feature f_1 as $f_1 = \operatorname{argmax}_f C(f)$.
- Select the second feature f_2 as:

$$f_2 = \operatorname{argmax}_{f \neq f_1} \{\alpha_1 C(f) - \alpha_2 |\rho_{f_1 f}|\},$$

where α_1 and α_2 are weighting factors.

- Select f_l , $l = 1, \dots, L$, such that:

$$f_l = \operatorname{argmax}_{f \neq f_r} \left\{ \alpha_1 C(f) - \frac{\alpha_2}{l-1} \sum_{r=1}^l |\rho_{f_r f}| \right\}, \quad r = 1, 2, \dots, l-1$$

6 Experiments

The approach described above has been implemented and tested in several 2D maps and 3D scenes. The goal of the experiment is to show the effectiveness of the iAMN in different indoor range data.

6.1 Classification of Places in 2D Maps

This experiment was carried out using the occupancy grid map of the building 79 at the University of Freiburg. For efficiency reasons we used a grid resolution of $20cm$, which lead us to a graph of 8088 nodes. The map was divided into two parts, the left one used for learning, and the right one used

for classification purposes (Figure 1). For each cell we calculate 203 geometrical features. This number was reduced to 30 applying the feature selection of Section 5. The right image of Figure 1 shows the resulting classification with a success rate of 97.6%.

6.2 Classification of Object in 3D Scenes

In this experiment we classify 3D scans of objects that appear in a laboratory of the building 79 of the University of Freiburg. The laboratory contain tables, chairs, monitors and ventilators. For each object class, an iAMN is trained with 3D range scans each containing just one object of this class (apart from tables, which may have screens standing on top of them). Figure 2 shows three example training objects. A complete laboratory in the building 79 of the University of Freiburg was later scanned with a 3D laser. In this 3D scene all the objects appear together and the scene is used as a test set. The resulting classification is shown in Figure 3. In this experiment 76.0% of the 3D points were classified correctly.

6.3 Comparison with Previous Approaches

In this section we compare our results with the ones obtained using other approaches for place and object classification. First, we compare the classification of the 2D map when using a classifier based on AdaBoost as shown by Martinez Mozos et al. (2005). In this case we obtained a classification rate of 92.1%, in contrast with the 97.6% obtained using iAMNs. We believe that the reason for this improvement is the neighboring relation between classes, which is ignored when using the AdaBoost approach. In a second experiment, we compare the resulting classification of the 3D scene with the one obtained when using AMN and NN. As we can see in Table 1, iAMNs perform better than the other approaches. A posterior statistical analysis using the *t*-student test indicates that the improvement is significant at the 0.05 level. We additionally realized different experiments in which we used the 3D scans of isolated objects for training and test purposes. The results are shown in Table 1 and they confirm that iAMN outperform the other methods.

Table 1. Classification results in 3D data

Data set	NN	AMN	iAMN
Complete scene	63%	62%	76%
Isolated objects	81%	72%	89%



Fig. 1. The left image depicts the training map of building 79 at the University of Freiburg. The right image shows the resulting classified map using an iAMN with 30 selected features.

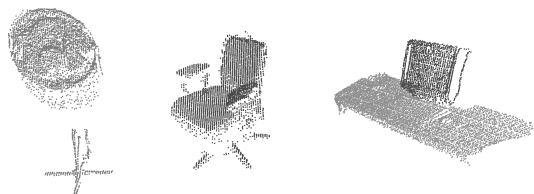


Fig. 2. 3D scans of isolated objects used for training: a ventilator, a chair and a table with a monitor on top.



Fig. 3. Classification of a complete 3D range scan obtained in a laboratory at the University of Freiburg.

7 Conclusions

In this paper we propose a semantic classification algorithm that combines associative Markov networks with an instance-based approach based on nearest neighbor. Furthermore, we show how this method can be used to classify points described by features extracted from 2D and 3D laser scans. Additionally, we present an approach to reduce the number of features needed to represent each data point, while maintaining their class discriminatory information. Experiments carried out in 2D and 3D maps demonstrated the

effectiveness of our approach for semantic classification of places and objects in indoor environments.

8 Acknowledgment

This work has been supported by the EU under the project CoSy with number FP6-004250-IP and under the project BACS with number FP6-IST-027140.

References

- ALTHAUS, P. and CHRISTENSEN, H.I. (2003): Behaviour Coordination in Structured Environments. *Advanced Robotics*, 17(7), 657–674.
- ANGUELOV, D., TASKAR, B., CHATALBASHEV, V., KOLLER, D., GUPTA, D., HEITZ, G. and NG, A. (2005): Discriminative Learning of Markov Random Fields for Segmentation of 3D Scan Data. *IEEE Computer Vision and Pattern Recognition*.
- BOYKOV, Y. and HUTTENLOCHER, D. P. (1999): A New Bayesian Approach to Object Recognition. *IEEE Computer Vision and Pattern Recognition*.
- FRIEDMAN, S., PASULA, S. and FOX, D. (2007): Voronoi Random Fields: Extracting the Topological Structure of Indoor Environments via Place Labeling. *International Joint Conference on Artificial Intelligence*.
- HUBER, D., KAPURIA, A., DONAMUKKALA, R. R. and HEBERT, M. (2004): Parts-Based 3D Object Classification. *IEEE Computer Vision and Pattern Recognition*.
- JOHNSON, A. (1997): *Spin-Images: A Representation for 3-D Surface Matching*. PhD thesis, Robotics Institute, Carnegie Mellon University, Pittsburgh, PA.
- KOENIG, S. and SIMMONS, R. (1998): Xavier: A Robot Navigation Architecture Based on Partially Observable Markov Decision Process Models. In: Kortenkamp, D. and Bonasso, R. and Murphy, R. (Eds). *Artificial Intelligence Based Mobile Robotics: Case Studies of Successful Robot Systems*. MIT-Press, 91–122.
- MARTINEZ MOZOS, O., STACHNISS, C. and BURGARD, W. (2005): Supervised Learning of Places from Range Data using Adaboost. *IEEE International Conference on Robotics & Automation*.
- MORAVEC, H. P. (1988): Sensor Fusion in Certainty Grids for Mobile Robots. *AI Magazine*, 61–74.
- OSADA, R., FUNKHOUSER, T., CHAZELLE, B. and DOBKIN, D. (2001): Matching 3D Models with Shape Distributions. *Shape Modeling International* 154–166.
- TASKAR, B., CHATALBASHEV, V. and KOLLER, D. (2004): Learning Associative Markov Networks. *International Conference on Machine Learning*.
- THEODORIDIS, S. and KOUTROUMBAS, K. (2006): *Pattern Recognition*. Academic Press, 3rd Edition, 2006.
- TRIEBEL, R., SCHMIDT, R., MARTINEZ MOZOS, O. and BURGARD, W. (2007): Instance-based AMN Classification for Improved Object Recognition in 2D and 3D Laser Range Data. *International Joint Conference on Artificial Intelligence*