Learning Semantic Maps with Topological Spatial Relations
Using Graph-Structured Sum-Product Networks

Kaiyu Zheng, Andrzej Pronobis and Rajesh P. N. Rao

Abstract

We introduce Graph-Structured Sum-Product Networks (GraphSPNs), a probabilistic approach to structured prediction for problems where dependencies between latent variables are expressed in terms of arbitrary, dynamic graphs. While many approaches to structured prediction place strict constraints on the interactions between inferred variables, many real-world problems can be only characterized using complex graph structures of varying size, often contaminated with noise when obtained from real data. Here, we focus on one such problem in the domain of robotics. We demonstrate how GraphSPNs can be used to bolster inference about semantic, conceptual place descriptions using noisy topological relations discovered by a robot exploring large-scale office spaces. Through experiments, we show that GraphSPNs consistently outperform the traditional approach based on undirected graphical models, successfully disambiguating information in global semantic maps built from uncertain, noisy local evidence. We further exploit the probabilistic nature of the model to infer marginal distributions over semantic descriptions of as yet unexplored places and detect spatial environment configurations that are novel and incongruent with the known evidence.

Introduction

It is essential for a mobile robot to maintain a representation of spatial knowledge, a framework that organizes the understanding of the environment. Mobile robots have access to information at both local and global scale. Therefore, it is desirable for a representation to enable integration of knowledge across spatial scales and levels of abstraction with the help of discovered spatial relations. Topological maps are an established framework for representing spatial relations between local places that enables anchoring high-level conceptual information and easy access for a planning algorithm. As a result, several semantic mapping approaches rely on topological graphs as part of their representation and associate topological nodes with semantic place attributes (Friedman, Pasula, and Fox 2007; Zender et al. 2008; Pronobis and Jensfelt 2012).

In order to integrate the collected spatial knowledge, resolve ambiguities, and make predictions about unobserved places, such frameworks often employ structured prediction algorithms. Unfortunately, the relations discovered by a robot exploring a real-world environment tend to be complex and noisy, resulting in difficult inference problems. Topological maps are dynamic structures, growing as the robot explores its environment, and containing a different number of nodes and relations for every environment. At the same time, many approaches to structured prediction place strict constraints on the interactions between inferred variables to achieve tractability (Bach and Jordan 2002) and require that the number of output latent variables be constant and related through a similar global structure (Belanger and McCallum 2016). This makes them either inapplicable or impractical in robotics settings and require compromising on the structure complexity (Mozos et al. 2007), introducing prior structural knowledge (Friedman, Pasula, and Fox 2007) or making hard commitments about values of semantic attributes (Pronobis and Jensfelt 2012). These problems are not unique to robotics and often present themselves in other domains, such as computer vision (Schwing and Urtasun 2015).

In this paper, we present Graph-Structured Sum-Product Networks (GraphSPNs), a general probabilistic framework for modeling graph-structured data with complex, noisy dependencies between a varying number of latent variables. Our framework builds on Sum-Product Networks (SPNs) (Poon and Domingos 2011), a probabilistic deep architecture with solid theoretical foundations (Peharz et al. 2017). SPNs can learn probabilistic models capable of representing context-specific independence directly from high-dimensional data and perform fast, tractable inference on high-treewidth models. GraphSPNs learn template SPN models representing distributions over attributes of subgraphs of arbitrary complexity. Then, to perform inference for a specific, potentially expanding graph, they assemble a mixture model over multiple decompositions of the graph into sub-graphs, with each sub-graph modeled by an instantiation of an appropriate template.

We apply GraphSPNs to the problem of modeling large-scale, global semantic maps with noisy topological spatial relations built by robots exploring multiple office environ-
ments. We make no assumptions about the structure of the topological map that would simplify the inference over semantic attributes. Our approach is capable of disambiguating uncertain and noisy local information about semantic attributes of places as well as inferring distributions over semantic attributes for yet unexplored places, for which local evidence is not available. Furthermore, we illustrate the benefits of the probabilistic representation by relying on likelihood of global semantic maps to detect novel and incongruent spatial configurations. We compare the performance of our model with the traditional approach based on Probabilistic Graphical Models assembled according to the structure of the topological graph. We show that GraphSPNs significantly outperforms Markov Random Fields built from pairwise and higher-order potentials relying on the same uncertain evidence.

**Related Work**

Probabilistic graphical models (PGMs) (Koller and Friedman 2009) provide a flexible framework for structured prediction. PGMs express the conditional dependence between random variables in terms of a graph. This results in a factorization of the joint distribution into a normalized products of factors, with factors defined over subsets of variables. The factors can be considered templates from which a distribution over an arbitrary number of variables can be assembled. As a result, PGMs were used in the past for modeling graph-structured data (Friedman, Pasula, and Fox 2007; Pronobis and Jensfelt 2012). Unfortunately, inference in PGMs is generally intractable, with the exception of low treewidth models (Bach and Jordan 2002). Even models with pairwise potentials require approximate inference techniques with no guarantee of convergence, such as Loopy Belief Propagation (BP) (Murphy, Weiss, and Jordan 1999), when the graph structure involves loops (as in this work). Higher-order models can pose a challenge even to the approximate methods. Additionally, many distributions which can be represented compactly, cannot be represented in terms of the factorization performed by PGMs (Poon and Domingos 2011).

An increasing number of structure prediction approaches utilize deep architectures. Several approaches build on graphical models for representing dependencies between output variables, while relying on deep models to realize the factors (Schwing and Urtasun 2015; Chen et al. 2015). Other types of models focus specifically on sequential data by combining local convolutions with classification (Collobert et al. 2011) or by building sequence to sequence predictors based on multiple LSTM layers (Sutskever, Vinyals, and Le 2014). Finally, fully convolutional models exist for such problems as pixel-wise segmentation (Shelhamer, Long, and Darrell 2017). Other deep models rely on deep architectures to define an energy function over candidate labels and then predictions are produced by using backpropagation to iteratively optimize the energy with respect to the labels. Structured Prediction Energy Networks (SPENs) (Belanger and McCallum 2016) take that approach and can capture dependencies that would lead to intractable graphical models. Unfortunately, many of the deep approaches are not probabilistic and are mostly applicable to data of the same global structure and number of output labels as the training examples.

Sum-Product Networks (SPNs) are unique in being a deep and probabilistic architecture capable of representing high-treewidth models that would result in unmanageable PGMs. Yet, they provide tractable exact inference by exploiting context-specific independence and determinism (Gens and Domingos 2012). In contrast to SPENs and approximate inference over PGMs, marginal and MPE inference in SPNs does not require iterations and can be achieved using a single pass through the network (Poon and Domingos 2011). SPNs are quite expressive and have been used to solve difficult problems in several domains (Poon and Domingos 2011; Gens and Domingos 2012; Amer and Todorovic 2015), including semantic place classification in robotics (Pronobis and Rao 2017). These benefits resulted in several approaches to structured prediction using SPNs. In (Ratajczak, Tschatschek, and Pernkopf 2014), Linear-chain Conditional Random Fields (LC-CRFs) extended with local factors modeled using SPNs were used to model sequences. In (Nath and Domingos 2015), Relational-SPNs were proposed that model graph-based relational data based on first-order logic. This method models graphs with potentially varying sizes by summarizing multiple variables with an aggregate statistic. In contrast, we directly model each output variable associated with nodes of the graph, and construct an SPN structure specific to each graph instance.

There have been numerous attempts to employ structured prediction to modeling semantic maps with topological spatial relations. In (Mozos et al. 2007), Hidden Markov Models were used to smooth sequences of AdaBoost classifications of place observations into semantic categories. (Friedman, Pasula, and Fox 2007) proposed Voronoi Random Fields (VRFs) which are CRFs constructed according to a Voronoi graph extracted from an occupancy grid map. VRFs utilize pairwise potentials to model dependency between neighboring graph nodes and 4-variable potentials to model junctions. In (Pronobis and Jensfelt 2012), Markov Random Fields were used to model pairwise dependencies between semantic categories of rooms according to a topological map. The categorical variables were connected to Bayesian Networks that reasoned about local environment features, forming a chain graph. This approach relied on a door detector to segment the environment into a topological graph with only one node per room. While these approaches are probabilistic, they rely on approximate inference using Loopy BP leading to problems with convergence (Friedman, Pasula, and Fox 2007). Moreover, in both cases, additional prior knowledge or hard commitments about the semantics of some places were employed in order to obtain a clean and manageable topological graph structure. In contrast, in this work, we rely on a graph built primarily to support navigation and execution of actions by the robot. Such graph provides a better coverage, but results in more noisy structure. Furthermore, we make no hard commitments about the semantics of the places at time of structure creation and defer such inference to the final model. Our experiments show, that under such conditions, graphical models with pairwise or higher-order potentials deteriorate quickly.
Figure 1: An simple SPN for a naive Bayes mixture model \( P(X_1, X_2) \), with three components over two binary variables. The bottom layer consists of indicators for \( X_1 \) and \( X_2 \). Weighted sum nodes, with weights attached to inputs, are marked with +, while product nodes are marked with \( \times \).

**Preliminaries**

We begin by giving a brief introduction to SPNs. For details, the reader is referred to (Peharz et al. 2017; Poon and Domingos 2011). Then, we describe the topological mapping framework we used to build the topological graphs.

**Sum-Product Networks**

One of the primary limitations of probabilistic graphical models is the complexity of their partition function, often requiring complex approximate inference in the presence of non-convex likelihood functions. In contrast, SPNs represent joint or conditional distributions with partition functions that are guaranteed to be tractable and involve a polynomial number of sum and product operations, permitting exact inference.

As shown in Fig. 1, an SPN is a directed acyclic graph composed of weighted sum and product operations. The sums can be seen as mixture models over subsets of variables, with weights representing mixture priors. Products can be viewed as combinations of features. SPNs can be defined for both continuous and discrete variables, with evidence for categorical variables often specified in terms of indicators. However, the latter can be obtained by replacing sum operations with weighted max operations (Peharz et al. 2017).

Parameters of an SPN can be learned generatively (Poon and Domingos 2011) or discriminatively (Gens and Domingos 2012) using Expectation Maximization (EM) or gradient descent. Additionally, several algorithms were proposed for simultaneous learning of network parameters and structure (Hsu, Kalra, and Poupart 2017; Gens and Domingos 2013). In this work, we use a simple structure learning technique (Pronobis and Rao 2017) to build template SPNs representing each sub-graph. We begin by initializing the SPN with dense structure by recursively generating nodes based on multiple random decompositions of the set of variables into multiple subsets until each subset is a singleton. The resulting structure consists of products combining the subsets in each decomposition and sums mixing different decompositions at each level. Then, we employ hard EM to learn the model parameters, which was shown to work well for generative learning (Pronobis and Rao 2017) and overcomes the diminishing gradient problem. After parameter learning, the generated structure can be pruned by removing edges associated with weights close to zero.

**Topological Graphs**

GraphSPNs are applicable to arbitrary graphs. However, in this work, we apply them specifically to topological graphs built by a mobile robot exploring a large-scale environment (Pronobis, Riccio, and Rao 2017). The primarily purpose of our topological graph is to support the behavior of the robot. As a result, nodes in the graph represent places the robot can visit and the edges represent navigability. The graph nodes are associated with latent variables representing semantics and the edges can be seen as spatial relations forming a global semantic map. Local evidence about the semantics of a place might be available and we assume that such evidence is inherently uncertain and noisy. Additional nodes in the graph are created to represent exploration frontiers, possible places the robot has not yet visited, but can navigate to. We call such nodes placeholders, and assume that the robot has not yet obtained any evidence about their semantics.

The topological graph is assembled incrementally based on dynamically expanding 2D occupancy map. The 2D map is built from laser range data captured by the robot using a
grid mapping approach based on Rao-Blackwellized particle filters (Grisetti, Stachniss, and Burgard 2007). Placeholders are added at neighboring, reachable, but unexplored locations and connected to existing places. Then, once the robot performs an exploration action, a placeholder is converted into a place and local evidence captured by the robot about the semantic place category is anchored to the graph node.

We formulate the problem of finding placeholder locations as sampling from a distribution that models location relevance and suitability. Specifically, the distribution is specified as: \( P(E|G, E) = \frac{1}{Z(G)} \sum_{S} \phi_{S}(E_{|G}) \phi_{N}(E_{|E}) \), where \( E_{i} \in \{0, 1\} \) represents the existence of a new place at location \( i \) in the occupancy map, \( G \) is the occupancy grid, and \( E \) is the set of locations of all existing places. The potential \( \phi_{S} \) ensures that placeholders are located in areas that are safe and preferred for navigation (are within safe distance from obstacles, with the preference towards centrally located places). The potential \( \phi_{N} \) models the neighborhood of a place, and guarantees sufficient coverage of the environment by promoting positions at a certain distance \( d_{m} \) from existing places. Final location of a new placeholder is chosen through MPE inference in \( P(E|G, E) \). An edge is then created to represent navigability. It connects the placeholder to an existing place in the graph based on A* search directly over the potential \( \phi_{S} \). An example of such semantic-topological map is shown in Fig. 4.

**GraphSPNs**

GraphSPNs learn a template model over arbitrary graph-structured data, with local evidence \( X_{i} \) and latent variables \( Y_{i} = \{Y_{i1}, \ldots, Y_{iM}\} \) for each graph node or edge \( i \), with dependencies between the latent variables expressed in terms of the graph structure. Then, an instance GraphSPN distribution \( P(X_{1}, Y_{1}, \ldots, X_{N}, Y_{N}) \) is assembled for a specific graph to perform inference.

To this end, we define a set \( S \) of template sub-graphs, and associate each template sub-graph \( S \in S \) with a separate template SPN modeling the distribution over variables \( X_{i} \) and \( Y_{i} \) corresponding to the nodes and edges of the template sub-graph. The structure and parameters of each template SPN can be learned directly from data obtained by decomposing training graphs into sub-graphs corresponding to \( S \).

Given a set of trained template SPNs, and a specific graph to be modeled, an instance GraphSPN is assembled as illustrated in Fig. 2. First, the graph is decomposed multiple times, each time differently, into sub-graphs using sub-graph templates \( S \) in descending order of the template size (i.e. more complex templates have priority). The sub-graphs should not overlap in each decomposition and the corresponding template SPNs should cover all variables \( X_{i}, Y_{i} \) in the model. This condition guarantees completeness and decomposability resulting in a valid instance GraphSPN. For each decomposition and each sub-graph, we instantiate the corresponding template SPN resulting in multiple SPNs sharing weights and structure. The instantiations for a single graph decomposition are combined with a product node and the product nodes for all decompositions become children of a root sum node realizing the complete mixture model.

In order to incorporate the latent variables \( Y_{ij} \), we include an intermediate layer of product nodes into the template SPNs. As shown in Fig. 2, each such product node combines an arbitrary distribution \( D_{ij}^{k}(X_{i}) \) with an indicator \( \lambda_{Y_{ij}=c_{ij}^{k}} \) for a specific value \( c_{ij}^{k} \) of \( Y_{ij} \). The template SPN built on top of the product nodes can be learned from data and the distributions \( D_{ij}^{k}(X_{i}) \) can be arbitrary, potentially also realized with an SPN with data-driven structure.

In our experiments, we assumed only one latent variable (semantic place category) \( Y_{i} \) per graph node \( i \), with \( \text{Val}(Y_{i}) = \{c_{i}^{1}, \ldots, c_{i}^{L}\} \), and we defined \( D_{ij}^{k}(X_{i}) \) for a single hypothetical binary observation \( x_{i} \), which we assumed to be observed:

\[
D_{ij}^{k}(X_{i}) = \begin{cases} \alpha_{ij}^{k} & X_{i} = x_{i} \\ 1 - \alpha_{ij}^{k} & X_{i} = \bar{x}_{i} \end{cases}
\]  

Such simplification allows us to thoroughly evaluate GraphSPNs for the problem of learning topological semantic maps by directly simulating hypothetical evidence about the semantic category of varying uncertainty and under various noise conditions. Furthermore, it allows us to compare GraphSPNs with Markov Random Fields using the same \( \alpha_{ij}^{k} \) as the value of local potentials, i.e. \( \phi_{i}(Y_{i} = c_{i}^{k}) = \alpha_{ij}^{k} \). The proposed approach naturally extends to the case where a more complex distribution is used to model semantic place categories based on robot observations, such as the SPN-based approach presented in (Pronobis and Rao 2017). Note, that we still learn the structure of the template SPNs built on top of distributions \( D_{ij}^{k}(X_{i}) \).
| NL | D_{\text{groundtruth}} | D_{\text{incorrect}} |
|----|------------------------|------------------------|
| 1  | 0.991 (+/-0.001)        | 0.0                    |
| 2  | 0.913 (+/-0.015)        | 0.085 (+/- 0.056)      |
| 3  | 0.720 (+/-0.040)        | 0.090 (+/- 0.061)      |
| 4  | 0.434 (+/-0.054)        | 0.092 (+/-0.062)       |
| 5  | 0.316 (+/-0.030)        | 0.154 (+/- 0.055)      |
| 6  | 0.154 (+/-0.021)        | 0.217 (+/-0.074)       |

Table 1: Noise levels used in our experiments.

**Experimental Procedure**

**Dataset**

The semantic maps with topological relations used in our experiments were obtained by deploying the topological mapping on sequences of laser range data and odometry captured by a mobile robot exploring multiple large-scale environments (Pronobis and Caputo 2009). The dataset contains 99 sequences, and as a result 99 topological graphs, captured on 11 floors of 3 buildings in different cities. We identified 10 semantic place classes that are common for all buildings (e.g. a corridor, a doorway, a 1-person office, see Fig. 4 for a complete list) and annotated each topological graph node with its groundtruth class.

Our goal in this work is to evaluate the ability of GraphSPNs to disambiguate semantic place classes despite noisy and uncertain local evidence by exploiting spatial relations captured in a noisy topological graph. Probabilistic place classification algorithms, such as the SPN-based approach in (Pronobis and Rao 2017) associate decisions based on local observations with probability estimates. However, the certainty of a decision can be low or the decision can be incorrect. In order to measure how sensitive the evaluated approaches are to such noise, we simulate local evidence attached to topological graph nodes by adding increasing noise to groundtruth information.

To this end, for each node \( i \) in each topological graph, we generated a local evidence distribution with values \( P(Y_i = c^k, \mathbf{X}_i = x_i) = \alpha_i^k \). For each graph, we first randomly selected 20% of all nodes for which the most likely local result should be incorrect. For those, we selected a random incorrect class to be associated with the highest probability value. Then, we randomized the value \( D_{\text{incorrect}} \), which is a difference between the highest probability and the probability of the groundtruth class, from a uniform distribution in a range depending on the noise level. For the remaining 80% of nodes, we ensured that the groundtruth class is associated with the highest probability. However, we simulated uncertainty by randomizing the value \( D_{\text{groundtruth}} \), which is a difference between the probability of the groundtruth class and the second highest probability. With these constraints, we used random values for the remaining likelihoods and made sure that each distribution is normalized. Intuitively, lower \( D_{\text{groundtruth}} \) indicates higher uncertainty and higher \( D_{\text{incorrect}} \) indicates stronger noise. The statistics of the values of \( D_{\text{incorrect}} \) and \( D_{\text{groundtruth}} \) for the final evidence at different noise levels are shown in Tab. 1.

**Learning GraphSPNs**

We learned GraphSPNs from a simple set of sub-graph templates shown in Fig. 3, matching from 1 to 5 nodes and simple edge configurations. We assumed that each node is associated with a single latent variable \( Y_i \) representing the semantic place class. For each template with at least 2 nodes, we learned a template SPN of specific structure and parameters from sub-graph examples in the training set in a supervised fashion (\( Y_i \) was set to the groundtruth). Each training graph was partitioned in 10 different ways to obtain sub-graphs. For a single-node template we simply assumed a uniform SPN. During testing, we built the instance GraphSPN based on 5 different graph decompositions (see Fig. 4). In our experiments, we always learned GraphSPN from all graphs from two buildings in the dataset and tested on graphs with different evidence noise levels from the remaining building. GraphSPNs are implemented using LibSPN (Pronobis, Ranganath, and Rao 2017).

**Constructing Markov Random Fields**

We compared GraphSPNs to a traditional approach based on MRFs structured according to the represented graph. The MRF was constructed from two types of potentials: potential \( \phi_i(Y_i = c^k) = \alpha_i^k \) used to provide local evidence, and potentials modeling latent variable dependencies. For the latter, we tried two models: using pairwise potentials for each pair of variables associated with connected nodes or defined over three variables for three connected nodes in any configuration. In each case, the potentials were obtained by generating co-occurrence statistics of variable values in the training graphs used for learning GraphSPNs. Inference in the MRF was performed using Loopy BP implemented in the libDAI library (Mooij 2010).

**Experimental Results**

We performed several experiments comparing the learned GraphSPN model to the MRF models with pairwise potentials (marked as MRF-2) and three-variable potentials (marked as MRF-3). First, we tasked all models with disambiguating noisy local evidence about semantic place class for places visited by the robot. For each topological graph in the test set, we performed marginal inference\(^1\), based on which we selected the final classification result \( \arg\max_k P(Y_i = c^k | \mathbf{X} = \mathbf{x}) \).

\(^1\)We experimented with MPE inference over all latent variables achieving inferior results with all models.
Figure 4: Visualization of the results for a graph from Freiburg at noise level 4. The top row shows: the semantic map with groundtruth semantic place classes; the 20% of nodes for which the most likely evidence indicates an incorrect class (black nodes); the semantic classes inferred by the GraphSPN and the MRF-2. The place classes are: 1-person office (1PO), 2-person office (2PO), bathroom (BA), corridor (CR), doorway (DW), kitchen (KT), laboratory (LAB), large office (LO), meeting room (MR), utility room (UT). The bottom row illustrates the 5 decompositions used when assembling the instance GraphSPN (different colors indicate different sub-graph templates applied).

The percentage of correctly classified nodes in the graph averaged over all test graphs is shown in Tab. 2 and a visualization of results for a single graph together with the decompositions used to build the instance GraphSPN is shown in Fig. 4. We evaluated all assignments of the three buildings into training and test sets as well as different noise levels listed in Tab. 1. Each test set consisted of 5 topological maps, each with 3 random sets of noisy local evidence resulting in 15 different test graphs. Since the local evidence for 20% of nodes in each graph indicates an incorrect class as the most likely one, only accuracy greater than 80% demonstrates that the model was able to recover from the noise using learned spatial relations. Lower accuracy suggests that the incorrect evidence was too strong or that the correct evidence was too uncertain to influence the semantic class of a place.

Analyzing the performance reported in Tab. 2, we see that pairwise MRF performs well when there is little noise in the evidence, however it deteriorates quickly with increasing noise levels. At the same time, GraphSPN generating robust results (accuracy greater than 80%) despite substantial noise and uncertainty. With substantial noise, approximate Loopy BP inference for MRF converges to a solution consisting primarily of the dominant class (the corridor). At the same time, we see that using higher-order potentials with MRF actually hurts performance.

In the second experiment, we tasked the models with inferring marginal distributions over the semantic classes of places not yet visited by the robot (placeholders) for which local evidence is unavailable. We used the same setup as in the previous experiment, with local evidence for explored places including noise and uncertainty. Examples of such marginal distributions are shown in Fig. 5, while the classification accuracy when considering the most likely class is reported in Tab. 3. Again, GraphSPN significantly outperformed the MRF for this inference task. If we analyze the marginal distributions for three representative placeholders shown in Fig. 5, we see that GraphSPN is confident about the correct class for the two placeholders for which nearby nodes provide correct (albeit uncertain) evidence. For the placeholder connected to nodes for which evidence indicates an incorrect class the marginal distribution is almost uniform. This indicates the ability of GraphSPN to generate useful confidence signal in presence of noisy inputs.

Our final experiment was designed to evaluate the quality of likelihood produced by GraphSPN instantiated over a graph of arbitrary size. In this experiment, we tasked the models with detecting novel spatial environment configurations that are incongruent with the evidence available during training. To this end, we produced novel graphs by swapping groundtruth semantic class labels between corridors.
Table 2: Semantic place classification accuracy for all models and test buildings, and at different noise levels.

|        | GraphSPN | MRF-2 | MRF-3 |
|--------|----------|-------|-------|
| NL Freiburg | Saarbrücken | Stockholm |     |
| 1 | 96.13%(+/-2.41) | 95.45%(+/-3.05) | 93.98%(+/-1.90) |
| 2 | 96.63%(+/-2.39) | 96.37%(+/-3.03) | 94.01%(+/-2.12) |
| 3 | 92.45%(+/-2.43) | 93.43%(+/-2.85) | 92.04%(+/-2.57) |
| 4 | 91.88%(+/-2.76) | 92.91%(+/-2.72) | 86.31%(+/-2.06) |
| 5 | 90.13%(+/-3.47) | 90.12%(+/-3.56) | 83.51%(+/-3.13) |
| 6 | 80.83%(+/-5.21) | 80.08%(+/-3.95) | 69.49%(+/-5.69) |

Table 3: Accuracy of semantic class inference for placeholders without local evidence, for all models and test buildings, and at two representative noise levels.

|        | GraphSPN | MRF-2 | MRF-3 |
|--------|----------|-------|-------|
| NL Freiburg | Saarbrücken | Stockholm |     |
| 2 | 67.58%(+/-10.42) | 78.15%(+/-9.95) | 67.57%(+/-11.11) |
| 5 | 40.59%(+/-12.22) | 55.18%(+/-19.67) | 37.56%(+/-10.44) |

Conclusions

We presented GraphSPNs, a probabilistic deep model for graph-structured data that learns a template distribution allowing for making inferences about graphs of different global structure. While existing works applied SPNs to data organized as fixed-size grids or sequences, this paper presents a novel attempt at deploying SPNs on arbitrary graphs of varying size. Based on GraphSPNs, we proposed a method for learning topological spatial relations in semantic maps constructed by a mobile robot. Our method is robust to noise and uncertainty inherent in real-world problems where information about the environment is captured with robot sensors. Our framework is universal and compatible with any distributions defined over local evidence. However, it is particularly well suited for integration with other task (as measured by the area under the curve). We see that GraphSPN performs comparably to the MRF models demonstrating its ability to produce useful likelihood values. At the same time, MRFs perform slightly better on this task. This is a result of lack of noise and uncertainty in the provided evidence. It shows that MRFs can capture relevant spatial relations and suggests that their performance deteriorates due to the limitations of approximate inference.
SPN-based models. In the future, by combining a Graph-SPN learned over semantic maps with the generative place model proposed in (Pronobis and Rao 2017), we intend to achieve a unified, deep, and hierarchical representation of spatial knowledge spanning from local sensory observations to global conceptual descriptions.

References

Amer, M., and Todorovic, S. 2015. Sum product networks for activity recognition. Transactions on Pattern Analysis and Machine Intelligence 38(4).

Bach, F. R., and Jordan, M. I. 2002. Thin junction trees. In Advances in Neural Information Processing Systems (NIPS).

Belanger, D., and McCallum, A. 2016. Structured prediction energy networks. In Proceedings of the International Conference on Machine Learning (ICML).

Chen, L.-C.; Schwing, A.; Yuille, A.; and Urtasun, R. 2015. Learning deep structured models. In Proceedings of the International Conference on Machine Learning (ICML).

Collobert, R.; Weston, J.; Bottou, L.; Karlen, M.; Kavukcuoglu, K.; and Kuksa, P. 2011. Natural language processing (almost) from scratch. Journal of Machine Learning Research 12(Aug).

Friedman, S.; Pasula, H.; and Fox, D. 2007. Voronoi random fields: Extracting the topological structure of indoor environments via place labeling. In Proceedings of the International Joint Conference on Artificial Intelligence (IJCAI).

Gens, R., and Domingos, P. 2012. Discriminative learning of sum-product networks. In Advances in Neural Information Processing Systems (NIPS).

Gens, R., and Domingos, P. 2013. Learning the structure of sum-product networks. In Proceedings of the International Conference on Machine Learning (ICML).

Grisetti, G.; Stachniss, C.; and Burgard, W. 2007. Improved techniques for grid mapping with Rao-Blackwellized particle filters. Transactions on Robotics 23(1).

Hsu, W.; Kalra, A.; and Poupart, P. 2017. Online structure learning for Sum-Product networks with gaussian leaves. arXiv:1701.05265.

Koller, D., and Friedman, N. 2009. Probabilistic Graphical Models: Principles and Techniques. The MIT Press.

Mooij, J. M. 2010. libDAI: A free and open source c++ library for discrete approximate inference in graphical models. Journal of Machine Learning Research 11.

Mozos, O. M.; Triebel, R.; Jensfelt, P.; Rottmann, A.; and Burgard, W. 2007. Supervised semantic labeling of places using information extracted from sensor data. Robotics and Autonomous Systems 55(5).

Murphy, K. P.; Weiss, Y.; and Jordan, M. I. 1999. Loopy belief propagation for approximate inference: An empirical study. In Proceedings of the Conference on Uncertainty in Artificial Intelligence (UAI).

Nath, A., and Domingos, P. 2015. Learning relational Sum-Product networks. In Proceedings of the AAAI Conference on Artificial Intelligence.